

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 15:49:42 ; Search time 436.05 Seconds
(without alignments)
145.869 Million cell updates/sec

Title: US-09-037-472-10

Perfect score: 20

Sequence: 1 TTACATATGAGCCTTCCATG 20

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl: *
1: gb_ba1: *
2: gb_ba2: *
3: gb_ov: *
4: gb_ov: *
5: gb_pat: *
6: gb_ph: *
7: gb_p11: *
8: gb_p12: *
9: gb_prl: *
10: gb_prl: *
11: gb_prl: *
12: gb_ro: *
13: gb_st: *
14: gb_sts: *
15: gb_sy: *
16: gb_un: *
17: gb_v1: *
18: em_fun: *
19: em_hgt: *
20: em_hum1: *
21: em_hum2: *
22: em_in: *
23: em_om: *
24: em_or: *
25: em_ov: *
26: em_pat: *
27: em_ph: *
28: em_pl: *
29: em_ro: *
30: em_sts: *
31: em_sy: *
32: em_un: *
33: em_v1: *
34: gb_htg1: *
35: gb_htg2: *
36: gb_in1: *
37: gb_in2: *
38: em_ba1: *
39: em_ba2: *
40: em_hum3: *
41: em_hum4: *
42: gb_prl: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	5	173226 Sequence 2

C 2	18.4	92.0	11970	10	HSILIAG	X03833 Human gene
C 3	16.8	84.0	118605	34	HSDJ925J7	AL078622 Homo sapi
C 4	16.4	82.0	6265	5	E05467	E05467 DNA sequenc
C 5	16.4	82.0	6042	7	SCYBR208C	Z36077 S.cerevisia
C 6	16.4	82.0	6141	8	YSCAMD	M6926 Yeast urea
C 7	16.4	82.0	175555	9	HS596H12	AL031347 Human DNA
C 8	16.4	82.0	120187	11	AC005884	AC005884 Homo sapi
C 9	16.4	82.0	973	12	CMU17165	U17165 Cricetus
C 10	15.8	79.0	450	2	AF098256	AF098256 Pseudomon
C 11	15.8	79.0	41175	7	SC41KCIIV	Z67750 S.cerevisia
C 12	15.8	79.0	1208	7	SCCDC36	X04287 Yeast cell
C 13	15.8	79.0	3849	7	SCCDC9	X03246 Yeast CDC9
C 14	15.8	79.0	944	7	SCYDL165W	Z74213 S.cerevisia
C 15	15.8	79.0	7008	8	SCSFAARP	X68020 S.cerevisia
C 16	15.8	79.0	39283	9	DB7014	DB7014 Human (lamb
C 17	15.8	79.0	363	9	HSACHS5	X02505 H.sapiens g
C 18	15.8	79.0	194523	10	HS323M22	AL022476 Human DNA
C 19	15.8	79.0	39716	10	HSAC000356	AC000356 Human cos
C 20	15.8	79.0	284	10	HSIGLV311	Z73645 H.sapiens I
C 21	15.8	79.0	39279	11	AC000027	AC000027 Homo sapi
C 22	15.8	79.0	58725	11	AC003988	AC003988 Human PAC
C 23	15.8	79.0	124293	11	AC006007	AC006007 Homo sapi
C 24	15.8	79.0	7138	17	HPEA	M80581 Hepatitis E
C 25	15.8	79.0	7207	17	HPECG	D11092 Hepatitis E
C 26	15.8	79.0	144027	34	HS106502	AL035562 Homo sapi
C 27	15.8	79.0	227175	35	AC007253	AC007253 Homo sapi
C 28	15.8	79.0	165729	35	AC007254	AC007254 Homo sapi
C 29	15.8	79.0	165062	35	AC007688	AC007688 Homo sapi
C 30	15.4	77.0	2714	4	XLLAMU3	X13169 Xenopus lae
C 31	15.4	77.0	87967	8	AC005223	AC005223 Arabidops
C 32	15.4	77.0	139350	9	HS443E24	AL021367 Human DNA
C 33	15.4	77.0	41369	10	CH19F15314	AL022728 Human DNA
C 34	15.4	77.0	26524	10	HS447D17	AL022728 Human DNA
C 35	15.4	77.0	1614	10	HSXISTA	X56199 Human XIST
C 36	15.4	77.0	16481	10	HUMXIST	M97168 Homo sapien
C 37	15.4	77.0	198582	11	AC005291	AC005291 Homo sapi
C 38	15.4	77.0	135886	11	HSXIST2	U80460 Human Xq13
C 39	15.4	77.0	144723	12	AC005818	AC005818 , complet
C 40	15.4	77.0	2620	12	MDSRP105	D37797 Mouse mRNA
C 41	15.4	77.0	124110	34	HSDJ93K22	AL050333 Homo sapi
C 42	15.4	77.0	170965	35	AC005302	AC005302 Mus muscu
C 43	15.2	76.0	10981	1	AE000848	AE000848 Methanoba
C 44	15.2	76.0	1999	1	SMB5BPB2X	Y10532 S.mitis bbp
C 45	15.2	76.0	160000	35	AC007140	AC007140 Homo sapi

ALIGNMENTS

RESULT	1					
I73226						
LOCUS	I73226	Sequence 2	20 bp	DNA	PAT	23-DEC-1997
DEFINITION	Sequence 2	from patent US 5686246.				
ACCESSION	I73226					
NID	93009365					
VERSION	I73226.1	Gr:3009365				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unclassified.					
REFERENCE	1 (bases 1 to 20)					
AUTHORS	Kornman,K.S. and Duff,G.W.					
TITLE	Detecting genetic predisposition to periodontal disease					
JOURNAL	Patent: US 5686246-A 2 11-NOV-1997;					
FEATURES	Location/Qualifiers					
source	1..20					
BASE COUNT	5 a	5 c	3 g	7 t		
ORIGIN						

Query Match 100.0%; Score 20; DB 5; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.22;


```
/db_xref="GI:536588"
/db_xref="SWISS-PROT:P32528"
/translation="MTVSDTAAEISLGSIOIDWIDFHKSSSSQASRLLESLLDSQN
VAPDNAMWISLISKENLHQFOLKSRNENKTLPLGVPIAVKNDIVDGLRTTAACP
SFAYEBSKSVKVELLRNAGALIVGKTNLDQFATGLVGRTPSGYKTPCAFSKEHVSGG
SSAGSASVARGIPLVIALGDTAGSRVPAALNNLGLKPTKGVSCQGVPAACKSLD
CVSIFALNLSDAEFCRIMCOPDNDSEYSPVSNPLKFFSSNVITAIAPKNIPWGE
TKNPVLFSNAVENLSTGANVIEIDFELLEARCLYEGTWAEQAIQSFELDKPP
QESDLPVTSIIIEGAKKYSAVDCSFSEYKRGQILQKVRLLSVDVLCVPTCPNPTM
KQVADPVLVNSRGQWTFNVLADLAALAVPAGRDGLPNGITLIGKFTDYALLE
LANRYQNFPLNGSRGTGNTSSVKPANDOLVGPDPSTSIKLVAVGAHLKGLPLH
WOLEKNATYLCCTTKTAKYOLFALPKNGPVLPCLRRVQDSNGSIOIELEVSVPEL
FGAISMPPELPGIGSVELESGEWIKSFICEESGYKAKGTVDITKGGRAIFEMELKK
KESQKKLFDFVLIANRGEIAVRIITKLKLGIRSAVSDPDQSHQVTFVGPSSGDI
HGTAAQTYLDNMKIIDAAKOTNAQAIIPGYGFLSENADFSDACTSAGITFVGPSSDI
IRGLGLKHSARQIAQKAGVPLVPGSLITSVEEAKKAAELEYPMVKSTAGGIGL
QKVDSEDIHIFETVKHOGTEFGDAGVFLERFENARHVEVQLMGDFGKAIALGE
RDSLQRNOKVIEETPAPNPEKTRALAKAAESLGLNLYKCACTVEFIYDEKDE
FYLEVNRLQVHEPITEMVTGLDLVEMIRIAANDAPDFDSTKRVNGVSMEARLYA
ENPLKNRPSGLLVDPFDMARVDTWVKGTNISPEYDPTLAKIIVHGKDRDAIS
KLNAQLEETKVYGCITNIDYLSKIIITSDFFAKAKVSTNLSNQYEPTAIEITLPGA
TSIQDPRGVYGRIGVPPSGPMDAYSFRLANRIVGNDYRTPAIEVLTGSPVIFHCE
TVIATGTALCTLDGQEIPOHKPVEVKGSTLSIGKLTSGCRAYLIGIGIDVPKYL
GVSTFTLGNVGYNGRVLKLDVLELPSNEENKSVCELPONIPOSLIPOISETKEWR
IGVTCGPHGSPDFKPSIEEFTSEKKKVVHNSRFGVRLIGPKPKWARSNGEGGHH
PSNTHDYVLSGAINFTGDEPVIITCDGPSLGGFVCOAVVPEAEMLKVGQVPGDSIQ
FVPLSYESSLSKESODVAIKSLDGTKLRLDSVLSLPSETPILAOMEKVNLSPKV
VYKQKELLAVLAYETEIOFDENKMTISNIRLPMFADFSLACVQRYOETIRSSA
PWLNNYDFIANYNGISREVDMLYSAREFVLGDLGDFLSPCAVLPDRHRLFGSK
YNSRFTYFERGAVGIGMYCIYAANSFGYOLVGRTPIDWKLCLAASSSEVFLMNP
FQVFEYFVSEEDLDKMTEDCNGYKVNIEKSVFEDHQEYLRWINAKDSIITAFQGGQ
LGERAEFAKLIQANSELKSVTKPDEEDFPEGAIEIVSEYSGRWKSIASVGDV
IEAGQGLLIIEAKMAEIIISAPKSGIKIKICHNGDMVDSGDIVAVIETLA"
BASE COUNT      1820 a 1313 c 1099 g 1810 t
ORIGIN
1 TTACATATGAGCCTTCCA 18
|||||
Db 5360 TTGCATATGAGCCTTCCA 5343
|||||
```

```
Query Match      82.0%; Score 16.4; DB 7; Length 6042;
Best Local Similarity 94.4%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TTACATATGAGCCTTCCA 18
|||||
Db 5360 TTGCATATGAGCCTTCCA 5343
|||||

RESULT 6
YSCUAMD 6141 bp DNA PLN 02-FEB-1999
LOCUS Yeast urea amidolyase (DURI.2) gene, complete cds.
DEFINITION M64926
ACCESSION M64926.1 GI:173121
NID 9173121
VERSION M64926.1 GI:173121
KEYWORDS urea amidolyase.
SOURCE Saccharomyces cerevisiae DNA.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 6141)
AUTHORS Genbauer,F.S. and Cooper,T.G.
TITLE The urea amidolyase (DURI.2) gene of Saccharomyces cerevisiae
JOURNAL DNA Seq. 2 (1), 19-32 (1991)
MEDLINE 92199240
FEATURES Location/Qualifiers
source 1..6141
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
gene 241..5748
/feature="DURI.2"
/db_xref="SGD:S0000412"
CDS 241..5748
/gene="DURI.2"
```

```
/EC_number="6.3.4.6"
/codon_start=1
/function="hydrolysis of urea to ammonia and CO2"
/evidence=experimental
/product="urea amidolyase"
/protein_id="AAC41643.1"
/db_xref="PIR:G173122"
/db_xref="GI:173122"
/translation="MTVSDTAAEISLGSIOIDWIDFHKSSSSQASRLLESLLDSQN
VAPDNAMWISLISKENLHQFOLKSRNENKTLPLGVPIAVKNDIVDGLRTTAACP
SFAYEBSKSVKVELLRNAGALIVGKTNLDQFATGLVGRTPSGYKTPCAFSKEHVSGG
SSAGSASVARGIPLVIALGDTAGSRVPAALNNLGLKPTKGVSCQGVPAACKSLD
CVSIFALNLSDAEFCRIMCOPDNDSEYSPVSNPKNFSSNVITAIAPKNIPWGE
TKNPVLFSNAVENLSTGANVIEIDFELLEARCLYEGTWAEQAIQSFELDKPP
KESDLPVTSIIIEGAKKYSAVDCSFSEYKRGQILQKVRLLSVDVLCVPTCPNPTM
QVADPVLVNSRGQWTFNVLADLAALAVPAGRDGLPNGITLIGKFTDYALLE
LANRYQNFPLNGSRGTGNTSSVKPANDOLVGPDPSTSIKLVAVGAHLKGLPLH
WOLEKNATYLCCTTKTAKYOLFALPKNGPVLPCLRRVQDSNGSIOIELEVSVPEL
FGAISMPPELPGIGSVELESGEWIKSFICEESGYKAKGTVDITKGVGRFAYFEMELKK
KESQKKLFDFVLIANRGEIAVRIITKLKLGIRSAVSDPDQSHQVTFVGPSSGDI
HGTAAQTYLDNMKIIDAAKOTNAQAIIPGYGFLSENADFSDACTSAGITFVGPSSDI
IRGLGLKHSARQIAQKAGVPLVPGSLITSVEEAKKAAELEYPMVKSTAGGIGL
QKVDSEDIHIFETVKHOGTEFGDAGVFLERFENARHVEVQLMGDFGKAIALGE
RDSLQRNOKVIEETPAPNPEKTRALAKAAESLGLNLYKCACTVEFIYDEKDE
FYLEVNRLQVHEPITEMVTGLDLVEMIRIAANDAPDFDSTKRVNGVSMEARLYA
ENPLKNRPSGLLVDPFDMARVDTWVKGTNISPEYDPTLAKIIVHGKDRDAIS
KLNAQLEETKVYGCITNIDYLSKIIITSDFFAKAKVSTNLSNQYEPTAIEITLPGA
TSIQDPRGVYGRIGVPPSGPMDAYSFRLANRIVGNDYRTPAIEVLTGSPVIFHCE
TVIATGTALCTLDGQEIPOHKPVEVKGSTLSIGKLTSGCRAYLIGIGIDVPKYL
GVSTFTLGNVGYNGRVLKLDVLELPSNEENKSVCELPONIPOSLIPOISETKEWR
IGVTCGPHGSPDFKPSIEEFTSEKKKVVHNSRFGVRLIGPKPKWARSNGEGGHH
PSNTHDYVLSGAINFTGDEPVIITCDGPSLGGFVCOAVVPEAEMLKVGQVPGDSIQ
FVPLSYESSLSKESODVAIKSLDGTKLRLDSVLSLPSETPILAOMEKVNLSPKV
VYKQKELLAVLAYETEIOFDENKMTISNIRLPMFADFSLACVQRYOETIRSSA
PWLNNYDFIANYNGISREVDMLYSAREFVLGDLGDFLSPCAVLPDRHRLFGSK
YNSRFTYFERGAVGIGMYCIYAANSFGYOLVGRTPIDWKLCLAASSSEVFLMNP
FQVFEYFVSEEDLDKMTEDCNGYKVNIEKSVFEDHQEYLRWINAKDSIITAFQGGQ
LGERAEFAKLIQANSELKSVTKPDEEDFPEGAIEIVSEYSGRWKSIASVGDV
IEAGQGLLIIEAKMAEIIISAPKSGIKIKICHNGDMVDSGDIVAVIETLA"
BASE COUNT      1838 a 1116 c 1335 g 1852 t
ORIGIN
1 TTACATATGAGCCTTCCA 18
|||||
Db 551 TTGCATATGAGCCTTCCA 568
|||||

Query Match      82.0%; Score 16.4; DB 8; Length 6141;
Best Local Similarity 94.4%; Pred. No. 35;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
QY 1 TTACATATGAGCCTTCCA 18
|||||
Db 551 TTGCATATGAGCCTTCCA 568
|||||

RESULT 7
HS596H12/c
LOCUS Human DNA sequence from clone 596H12 on chromosome 6p22.2-22.3.
DEFINITION Contains ESTs, STSs and GSSs, genomic marker D6S461 and TC, CA and
TA repeat polymorphisms, complete sequence.
ACCESSION AL031347
NID 94160208
VERSION HTG: CA repeat polymorphism; D6S461; TA repeat polymorphism; TC
KEYWORDS repeat polymorphism.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 175555)
AUTHORS Peck,A.
TITLE Direct Submission
JOURNAL Submitted (23-DEC-1998) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA. UK. E-mail enquiries: humquery@sanger.ac.uk Clone
```

requests: clonerequest@sanger.ac.uk
 On Jan 16, 1999 this sequence version replaced gi.3927943.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence is the entire insert of clone 596H12. This sequence
 has been finished according to sequence map criteria as follows. An
 attempt is made to resolve all sequencing problems, such as
 compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 This sequence was generated from part of bacterial clone contigs of
 human chromosome 6, constructed by the Sanger Centre Chromosome 6
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 596H12 is from the library RPC14 constructed at the Roswell Park
 Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

FEATURES
 Location/Qualifiers
 1. .175555
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosomes="6"
 /clone="596H12"
 /map="p22.2-22.3"
 /clone_lib="RPC14"
 710. .840
 /note="Tigger3(Golem) repeat: matches 2. .142 of consensus"
 838. .1095
 /note="Tigger3(Golem) repeat: matches 2781. .3038 of
 consensus"
 1540. .1928
 /note="LTR16A repeat: matches 18. .440 of consensus"
 2320. .2825
 /note="HERV16 repeat: matches 328. .940 of consensus"
 2910. .3374
 /note="LIP2 repeat: matches 5669. .6146 of consensus"
 3787. .4161
 /note="LIP7 repeat: matches 5771. .6142 of consensus"
 4167. .4545
 /note="HERV16 repeat: matches 1571. .1944 of consensus"
 6293. .6503
 /note="HERV16 repeat: matches 4443. .4661 of consensus"
 7087. .7374
 /note="AluJb repeat: matches 1. .289 of consensus"
 7492. .7746
 /note="LTR16A repeat: matches 25. .291 of consensus"
 7913. .8039
 /note="LTR16A repeat: matches 317. .442 of consensus"
 8762. .8831
 /note="MSTA repeat: matches 361. .426 of consensus"
 8832. .9142
 /note="AluSq repeat: matches 1. .313 of consensus"
 9143. .9475
 /note="MSTA repeat: matches 1. .361 of consensus"
 9222. .9706
 /note="match: GSS AQ219540"
 9884. .10190
 /note="AluSq repeat: matches 1. .308 of consensus"
 complement(11134. .11527)
 /note="match: GSS AQ086021"
 complement(11525. .11848)
 /note="match: GSS AQ062320"
 12210. .12442
 /note="MIR repeat: matches 8. .262 of consensus"
 12780. .13027
 /note="LIME repeat: matches 5333. .5788 of consensus"
 14101. .14503
 /note="L2 repeat: matches 972. .1381 of consensus"

repeat_region
 14703. .15322
 /note="L2 repeat: matches 2041. .2745 of consensus"
 15692. .15922
 /note="MSTB repeat: matches 234. .426 of consensus"
 15923. .16196
 /note="AluSq repeat: matches 1. .276 of consensus"
 16197. .16416
 /note="MSTB repeat: matches 2. .234 of consensus"
 complement(17504. .175317)
 /note="match: EST R02441"
 17689. .17822
 /note="MIR repeat: matches 123. .256 of consensus"
 19121. .19677
 /note="L1MC/D repeat: matches 5127. .5673 of consensus"
 complement(19601. .19957)
 /note="match: STS Z24329; genomic marker D6S461;
 polymorphic TC, CA and TA repeats in 19687. .19853"
 19686. .19835
 /note="75 copies 2 mer gt 82% conserved"
 20167. .20548
 /note="L1MEL repeat: matches 5774. .6151 of consensus"
 20989. .21102
 /note="MER81 repeat: matches 1. .114 of consensus"
 21255. .21347
 /note="MIR repeat: matches 168. .261 of consensus"
 22169. .22944
 /note="L2 repeat: matches 1898. .2746 of consensus"
 23048. .23230
 /note="MIR repeat: matches 29. .214 of consensus"
 23563. .23740
 /note="MER91A repeat: matches 1. .183 of consensus"
 24960. .25161
 /note="MER2 repeat: matches 8. .210 of consensus"
 complement(25042. .25379)
 /note="match: GSS AQ191276"
 25473. .25759
 /note="AluXc repeat: matches 1. .287 of consensus"
 26467. .26540
 /note="L2 repeat: matches 2237. .2321 of consensus"
 26573. .26786
 /note="L2 repeat: matches 2314. .2522 of consensus"
 27198. .27348
 /note="MLT1J repeat: matches 332. .496 of consensus"
 28127. .28410
 /note="AluSq repeat: matches 1. .286 of consensus"
 28995. .29065
 /note="MSTB repeat: matches 355. .426 of consensus"
 29066. .29361
 /note="AluY repeat: matches 1. .296 of consensus"
 29362. .29723
 /note="MSTB repeat: matches 1. .355 of consensus"
 31269. .31391
 /note="L2 repeat: matches 2638. .2750 of consensus"
 33043. .33380
 /note="MER57B repeat: matches 64. .400 of consensus"
 33076. .33435
 /note="MER93 repeat: matches 1. .361 of consensus"
 33945. .34322
 /note="match: GSS B87850"
 35631. .35737
 /note="L2 repeat: matches 2632. .2738 of consensus"
 complement(36270. .36631)
 /note="match: GSS AQ182323"
 37479. .37644
 /note="MLT1J repeat: matches 22. .181 of consensus"
 37689. .37776
 /note="MER5A repeat: matches 4. .87 of consensus"
 37898. .38022
 /note="MER5A repeat: matches 55. .189 of consensus"
 38042. .38099
 /note="MER5B repeat: matches 120. .173 of consensus"
 38100. .38827
 /note="LTR8 repeat: matches 1. .691 of consensus"

COMMENT

```

repeat_region 38828. .38873
/note="MER5B repeat: matches 62. .120 of consensus"
repeat_region 40174. .40709
/note="MER41C repeat: matches 9. .554 of consensus"
repeat_region 40720. .40879
/note="LTR16B repeat: matches 137. .304 of consensus"
repeat_region 40927. .41658
/note="MER50 repeat: matches 1. .722 of consensus"
repeat_region 42300. .42409
/note="MER5B repeat: matches 69. .178 of consensus"
repeat_region 42536. .43325
/note="L1MC1 repeat: matches 5555. .6325 of consensus"
repeat_region 43982. .44194
/note="MIR repeat: matches 28. .262 of consensus"
repeat_region 45047. .45203
/note="FRAM repeat: matches 1. .167 of consensus"
repeat_region 45278. .45477
/note="L1MC4 repeat: matches 7308. .7511 of consensus"
repeat_region 45478. .45843
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 45844. .46192
/note="THE1C-internal repeat: matches 1. .350 of consensus"
repeat_region 46198. .47354
/note="THE1C-internal repeat: matches 427. .1580 of consensus"
repeat_region 47355. .47733
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 47817. .48057
/note="L1MD3 repeat: matches 7484. .7739 of consensus"
repeat_region 49840. .49883
/note="22 copies 2 mer gt 93% conserved"
repeat_region 51669. .51763
/note="L2 repeat: matches 2628. .2708 of consensus"
repeat_region 52271. .52350
/note="L1MCE repeat: matches 1593. .1674 of consensus"
repeat_region 52472. .52566

Query Match 82.0%; Score 16.4; DB 9; Length 175555;
Best Local Similarity 94.4%; Pred. No. 42;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TACATATGAGCTTCAT 19
|||||
Db 54844 TACATATGAGCTTCAT 54827

RESULT 8
AC005884/C
LOCUS Homo sapiens chromosome 17, clone hRPK.264_B_14, complete
DEFINITION sequence.
ACCESSION AC005884
NID 93858902
VERSION AC005884.1 GI:3858902
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 120187)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 17, clone hRPK.264_B_14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 120187)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.

```

TITLE

JOURNAL

REFERENCE

AUTHORS

Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.

Submitted (28-OCT-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 120187)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stillwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M.

Submitted (10-NOV-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Nov 10, 1998 this sequence version replaced gi:3850604.

All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Only the last 120187 base pairs of this clone are being submitted.

The remainder overlaps with accession number AC005856(WICGR project L452).

FEATURES

source

Location/Qualifiers

1. .120187

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone=" hRPK.264_B_14"

/map="17"

/chromosome="17"

136. .253

/rpt_family="MIR"

1699. .1783

/rpt_family="MIR"

complement(1997. .2070)

/rpt_family="L2"

complement(3529. .3821)

/rpt_family="AluY"

4933. .5011

/rpt_family="MIR"

5197. .5497

/rpt_family="AluY"

complement(5510. .5833)

/rpt_family="L1M4C"

5844. .6096

/rpt_family="L1M3"

complement(6097. .6311)

/rpt_family="AluJb"

6312. .6497

/rpt_family="L1M3"

6524. .6663

/rpt_family="AluSg/x"

6671. .6776

/rpt_family="L1M3"

complement(6792. .6820)

/rpt_family="(CA)n"

complement(6888. .6898)

/rpt_family="Charliel"

6899. .7192

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```
repeat_region /rpt_family="AluSx"
complement(7193..7810)
repeat_region /rpt_family="Charliel"
complement(7811..8099)
repeat_region /rpt_family="AluSp"
complement(8100..8738)
repeat_region /rpt_family="Charliel"
8762..9063
/rpt_family="LIME3"
complement(9064..9357)
repeat_region /rpt_family="AluJb"
9358..9457
/rpt_family="LIME3"
9458..9765
/rpt_family="AluSx"
9766..10787
/rpt_family="LIME3"
11711..11990
/rpt_family="AluSg"
11991..12016
/rpt_family="(GAAA)n"
12834..13147
/rpt_family="AluSx"
complement(14452..14480)
repeat_region /rpt_family="(CAAAA)n"
complement(14481..14764)
repeat_region /rpt_family="AluSx"
16273..16576
/rpt_family="AluSx"
18854..19070
/rpt_family="AluSc"
complement(19232..19308)
repeat_region /rpt_family="L2"
20138..20305
/rpt_family="L1MC5"
20314..20446
/rpt_family="L1MB6"
complement(21891..22202)
repeat_region /rpt_family="AluJb"
complement(22235..22460)
repeat_region /rpt_family="L2"
22461..22782
/rpt_family="AluJo"
22785..23053
/rpt_family="AluSx"
23054..23085
/rpt_family="(TAAAA)n"
complement(23086..23144)
repeat_region /rpt_family="L2"
24563..24608
/rpt_family="(CA)n"
24876..24901
/rpt_family="AT_rich"
complement(25113..25147)
repeat_region /rpt_family="(TAAAA)n"
complement(25148..25429)
repeat_region /rpt_family="AluY"
25493..25536
/rpt_family="AT_rich"
26609..26787
/rpt_family="AluJo"
27051..27337
/rpt_family="AluSg"
27426..27478
/rpt_family="(TA)n"
complement(28432..28519)
repeat_region /rpt_family="L1MC1"
complement(28533..28837)
repeat_region /rpt_family="AluJo"
complement(29747..29960)
repeat_region /rpt_family="L2"
30103..30177
/rpt_family="L2"

repeat_region 30257..30440
/rpt_family="L1MC5"
31072..31236
/rpt_family="MERSA"
complement(31697..31726)
repeat_region /rpt_family="(CAA)n"
32265..32298
/rpt_family="(TA)n"
complement(32810..33109)
repeat_region /rpt_family="AluSx"
33695..33716
/rpt_family="AT_rich"
34886..34923
/rpt_family="(TA)n"
complement(34924..35228)
repeat_region /rpt_family="AluSc"
35229..35260
/rpt_family="(TA)n"
35344..35409
/rpt_family="(TA)n"
complement(38342..38515)
repeat_region /rpt_family="FAM"
38608..38635
/rpt_family="HSMAR2"
38632..39098
/rpt_family="HSMAR2"
39099..39409
/rpt_family="AluY"
39410..40054

Query Match 82.0%; Score 16.4; DB 11; Length 120187;
Best Local Similarity 94.4%; Pred No. 41;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCA 18
|||||
Db 74248 TTACTTATGAGCCTTCCA 74231

RESULT 9
CMU17165/c
LOCUS CMU17165 973 bp mRNA ROD 04-DEC-1995
DEFINITION Cricetulus migratorius Ig kappa light chain mRNA, complete cds.
ACCESSION U17165
NID 9841147
VERSION U17165.1 GI:8411147
KEYWORDS
SOURCE Armenian hamster.
ORGANISM Cricetulus migratorius
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE 1 (bases 1 to 766)
AUTHORS Whitters,M.J. and Collins,M.
TITLE Hamster cDNA homologs to the mouse immunoglobulin kappa constant
and Igh-V 45.1 genes
JOURNAL Immunogenetics 42 (3), 227-228 (1995)
MEDLINE 95369852
REFERENCE 2 (bases 1 to 973)
AUTHORS Collins,M.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-1994) Mary Collins, Genetics Institute, 87
Cambridge Park Dr., Cambridge, MA 02140, USA
FEATURES
Source
1..973
Location/Qualifiers
/organism="Cricetulus migratorius"
/db_xref="taxon:10032"
/dev_stage="rearranged"
50..769
/codon_start=1
/product="immunoglobulin kappa light chain"
/protein_id="AAA82732.1"
/db_xref="PID:g841148"
/db_xref="GI:841148"
```

/translation="MKLPVLLALLFMIPDSRGDVVMTQSPNVLSVSLGEQVYSICR
SSOSLVOSNNTVNNFLOPQSGPKRLIYKTSNRNSGVDPDTFSGSGSKDFTLKISR
MEYEDGVYCMGSSYPWTFEGTKLEIKRADAKETVSIFFPSSQOLGTCGSAIYVCF
VNNFVADINVKWVGDEARQVLOSVDQDSKSTYSLSSTLSUTKADYERHNDYT
CEVTHKTSTAIAIVKTLNRNEC"
50. .391
V_region
J_segment 392. .448
C_region 449. .766
BASE COUNT 267 a 219 g 241 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 12; Length 973;
Best Local Similarity 94.4%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 3 ACATATGAGCCTTCATG 20
|||||
Db 405 ACATATGAGCCTTCATG 388
|||||

RESULT 10
AF098256
LOCUS
DEFINITION Pseudomonas syringae pv. pisi strain 539 internal transcribed
spacer 1, complete sequence.
ACCESSION AF098256
NID 93873271
VERSION AF098256.1 GI:3873271
KEYWORDS
SOURCE
ORGANISM Pseudomonas syringae pv. pisi.
Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
Pseudomonas.

REFERENCE 1 (bases 1 to 450)
AUTHORS Koike,S.T., Barak,J.D., Henderson,D.M. and Gilbertson,R.L.
TITLE Bacterial blight of leek: A new disease in California caused by
Pseudomonas syringae
JOURNAL Plant Dis. (1998) In press
REFERENCE 2 (bases 1 to 450)
AUTHORS Barak,J.D., Gilbertson,R.L. and Koike,S.T.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Plant Pathology, University of California,
Davis, One Shields Blvd., Davis, CA 95616, USA

FEATURES
source
1. .450
/organism="Pseudomonas syringae pv. pisi"
/strain="539"
/db_xref="taxon:59510"
/note="pathovar: pisi"
misc_RNA
1. .450
/note="ITS1"
/product="internal transcribed spacer 1"
BASE COUNT 106 a 110 c 113 g 121 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 2; Length 450;
Best Local Similarity 89.5%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 TTACATATGAGCCTTCAT 19
|||||
Db 324 TTACAAATGAGCCTTCAT 342
|||||

RESULT 11
SC41KCIV
LOCUS
DEFINITION S. cerevisiae DNA (cosmid 31A2; chromosome IV; 41 kb).
ACCESSION 267750
NID g1061256
VERSION 267750.1 GI:1061256

KEYWORDS

SOURCE
ORGANISM

REFERENCE
AUTHORS
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

1. .41175
/organism="Saccharomyces cerevisiae"
/strain="alphaS288C"
/db_xref="taxon:4932"
/chromosome="IV"
/clone_lib="cosmid 31A2;"
/clone_lib="cosmid library (C. Jacq, ENS, France)"
<1. .766
/codon_start=2
/product="putative protein"
/protein_id="CAA91564.1"
/db_xref="PID:g1061257"
/db_xref="GI:1061257"
/db_xref="SWISS-PROT:P48570"
/translation="IGVNRVGIADTVGCANPROVYELIRTLKSVSCDIECHTHNDTG
CAIANAYTALEGARLDIVSLVIGERNGITPLGLMARMIVAADPVYKSKYLHKIR
DIENLVADAEVNPFPNPTGFCATFKAGIHAKAILANPSEIYELDPHI-GMKRVI
HFANRLTGWNAIKARVDQLNLTDDQIKEVTAKIKKLGVDRLNDDVDSIIKNFHA
EYSTPOVLSAKKNKNDSDVELATIPAAKRTKPSA"
1939. .3582
/codon_start=1
/product="putative protein"
/protein_id="CAA91565.1"
/db_xref="PID:g1061258"
/db_xref="GI:1061258"
/db_xref="SPTREMBL:Q12301"
/translation="MVRLNHAASYFMPICSTRPHIVISALFSISLSFYASSELL
LHOYDPLMKPNQSDYFRTFLGLSPFLYFKTLFNFNORFLITNLIVDFPIND
VFMLILIGLAYQVQDHGGTIKKKCSWHIIPROYIFGISWALGEFTICIGNLIF
NYQEIADPNINGFTHOESANTYCNNDMSHNDCCGSTEYRPNVDRSDITLSKIE
VRNDSSISNNVSYSEYHPKPLRSSSTYGSIRQPHENKKOLHVPDNDQDITIMM
NFDNSKLKTLTDLGSLPIDEEQPILKRSFGYTWAVPNTONTTKSTPIKRFIA
FSTAYOLVTGLLMLVGSNIMLTIGESLILSMYFVYVRGHEGLTPVNVYFGSRTI
SNFICVITISLNFILNTISYLRLEDLDFWNNQGEFEDDDENTISKRVATNQEYO
HPLSANYISMDSPDVINSFPGHGMNSGOLLGNTLYYGSGLNGDDDDMTNDSALLRRC
KKLVKNRALARNDSPVLGVMSVSLLVFVTGLSTVYI"
4334. .5248
/codon_start=1
/product="putative protein"
/protein_id="CAA91566.1"
/db_xref="PID:g1061259"
/db_xref="GI:1061259"
/db_xref="SPTREMBL:Q12477"
/translation="MISDIALQFNKKPVSQEMIOFLATSTASIIKIRNNNPICOC
RPPDJSIFKNNVIOGNVQTPMLATSVYLNKLKSVIPKNYVINTRRHIFLGLCIL
AAKTLNDSPNKKHWTYTEGLLRREVNTIRELLEYLNWDVRIITPDILDSLSYFL
GPIKQFLQRRQEMLLFNAPSPQOLKEYINHRRPVSHRSSTSAISVPISLTSNATYST
TDSRSSLLAKYQPSLPVSDNFNKKHVPRLNNDICNNFRAEENHSVNHIDVTMG
SSPVMSHKPTIHORLNFTRRGWSSEFKQ"
5365. .7157
/codon_start=1
/product="putative protein"
/protein_id="CAA91567.1"
/db_xref="PID:g1061260"
/db_xref="GI:1061260"
/db_xref="SWISS-PROT:P46681"
/translation="MLRNILVRSFGSNFKFAGRYMKSSALLGYRRVNYSTKIOTRL

CDS

CDS

CDS

CDS

TITLE Nucleotide sequence of the yeast cell division cycle start genes
CDC28, CDC36, CDC37, and CDC39, and a structural analysis of the
predicted products

JOURNAL Nucleic Acids Res. 14 (16), 6681-6697 (1986)

MEDLINE 86312926

FEATURES

Source Location/Qualifiers

gene 1..1208
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"

misc_feature 1..1208
/gene="CDC36"
/db_xref="SGD:S0002324"

misc_feature 24..29
/gene="CDC36"
/note="TATA-like sequence"
/db_xref="SGD:S0002324"

misc_feature 77..81
/gene="CDC36"
/note="TATA-like sequence"
/db_xref="SGD:S0002324"

misc_feature 145..147
/gene="CDC36"
/note="TATA-like sequence"
/db_xref="SGD:S0002324"

misc_feature 157..160
/gene="CDC36"
/note="TATA-like sequence"
/db_xref="SGD:S0002324"

misc_feature 189..192
/gene="CDC36"
/note="TATA-like sequence"
/db_xref="SGD:S0002324"

CDS 207..782
/gene="CDC36"
/note="CDC36"
/codon_start=1
/db_xref="SGD:S0002324"
/protein_id="CAA27835.1"
/db_xref="PID:g3491"
/db_xref="GI:3491"
/db_xref="SWISS-PROT:P06100"
/translation="MEKFLKALVPLKLEKLSSTYDHSMTLGLADLSMLYSIGIP
RDSQDHRVLDTFQSPWAEYSRSEPFPTPEFTNIPGVLSQSTVTPCFNSIONDQ
RVAFQDETFLFKYKHPGTVIQELTYLELRKWRHKTUKANLTKDPMPEIVSAD
GLSERGSYVFDPORWEKQCRDFFLFINAIM"

misc_feature 806..814
/gene="CDC36"
/note="pot. polyA signal"
/db_xref="SGD:S0002324"

misc_feature 822..826
/gene="CDC36"
/note="pot. termination sequence 1"
/db_xref="SGD:S0002324"

misc_feature 836..839
/gene="CDC36"
/note="pot. termination sequence 1"
/db_xref="SGD:S0002324"

misc_feature 893..896
/gene="CDC36"
/note="pot. termination sequence 2"
/db_xref="SGD:S0002324"

misc_feature 904..906
/gene="CDC36"
/note="pot. termination sequence 2"
/db_xref="SGD:S0002324"

BASE COUNT 390 a 226 c 240 g 352 t

ORIGIN

Query Match 79.0%; Score 15.8; DB 7; Length 1208;
Best Local Similarity 89.5%; Pred. NO. 73;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACATATGAGCCTTCATG 20
||||||| | |||||

Db 272 TACATATGACCATTCATG 290

RESULT 13

SCDC9/c 3849 bp DNA 10-FEB-1999

LOCUS Yeast CDC9 gene for DNA ligase.

DEFINITION X03246

ACCESSION X03246

NID g3514

VERSION X03246.1 GI:3514

KEYWORDS DNA ligase; inverted repeat; unidentified reading frame.

SOURCE baker's yeast.

ORGANISM Saccharomyces cerevisiae

REFERENCE 1 (bases 1 to 3849)

AUTHORS Barker,D.G., White,J.H. and Johnston,L.H.

TITLE The nucleotide sequence of the DNA ligase gene (CDC9) from Saccharomyces cerevisiae: a gene which is cell-cycle regulated and induced in response to DNA damage

JOURNAL Nucleic Acids Res. 13 (23), 8323-8337 (1985)

MEDLINE 86093646

COMMENT The direct repeat elements (1) may be implicated in the cell division cycle regulated expression of CDC9, 46 amino acids (pos. 2367-2504) show homology to regions of the T4 and T7 bacteriophage DNA ligases, and include the putative ATP binding site. CDC36 is included within the category of so-called 'start genes', encoding proteins which are required in early G1, when the cell is faced with the option of initiating a further cell cycle.

FEATURES

source Location/Qualifiers

1..3849
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"

gene 1..3849
/gene="CDC36"
/db_xref="SGD:S0002324"

gene 1..3849
/gene="CDC9"
/db_xref="SGD:S0002323"

repeat_region 86..91
/note="direct repeat 1"

repeat_region 114..119
/note="direct repeat 1"

repeat_region 157..162
/note="direct repeat 1"

repeat_region 251..256
/note="direct repeat 1"

repeat_region 369..374
/note="direct repeat 1"

repeat_region 568..582
/gene="CDC36"
/note="imp. inverted repeat A"

misc_signal 568..601
/db_xref="SGD:S0002324"

1..601
/gene="CDC36"

1..601
/note="pot. stem-loop structure pot. transcription regulator"

repeat_unit 587..601
/db_xref="SGD:S0002324"

repeat_region 661..666
/note="imp. inverted repeat A"

misc_feature 661..666
/db_xref="SGD:S0002324"

671..673
/gene="CDC36"

671..673
/note="major mRNA 5' end"

711..7978
/db_xref="SGD:S0002324"

CDS 711..7978
/gene="CDC36"
/note="put. DNA ligase (CDC9) (aa 1-755)";


```
/codon_start=1
/db_xref="SGD:S0002324"
/protein_id="CAA27005.1"
/db_xref="PID:g3515"
/db_xref="GI:3515"
/db_xref="SWISS-PROT:P04819"
/translation="MRLLTGCLSSARPLKSRLLPLMLSSSLPSSAGKKPKQATLARF
FTSMNKTEGTPSPKSKSKHLEMDNVSGEEYATKKKOTATVTHVAAPSSMG
NESSIPSSAPSGVADSPQSORLVGEVEDALSNNDHYSSNIPIYSEVCEVFNKIEA
ISSRLIEIRICSDFFIKMKQSSKNVIPTTYFINRLGPDYAGLEGLGELGNLLMKTI
STCSQKSLKIKYDKIDGEGEIAMGARNVOPTMFKPKPLTVGEVFNKLRATQ
KDSQKKKLKRLMTACKIEAKELIRLSLEKLRIGLAERTVLISLAKLLHDENR
EDSPKQPMVLESAQOKIRDAQCPVNYEIVINSCLHGMNLDKCYCTLRPGIPLK
PMLAKPTAINLEVDORFOGETFTSEYKYDGERAOVHLLNDGMTRIYSRNGENMTERY
EJNITDFODLDTNNLLDCEAVADKDGKILPEQVILSTRKDKDVELNDVKVKCL
FADILCYNDRLINKSLKERREYLTVKYVPGEGEQVATQITNNLDELQFLDES
NHSCEGLMYKLEGPESHYEPSKRRNMLKKDYELGVGSDLCVLGAYVGRKRT
GYGFLGCGYNODTGEFETCKIGTGFSDMLQLLHDLRTPTIIDGPKATVFDSSA
EPDWFPEPTTLFEVLTAJLSLPIYKAGSATFDKGVSLRFPRLRIRKDGVEDATSS
DOIVELYENQSHMQN"
complement(3101..3676)
/gene="CDC36"
/note="open reading frame (CDC36) (aa 1-191)"
/codon_start=1
/db_xref="SGD:S0002324"
/protein_id="CAA27006.1"
/db_xref="PID:g3516"
/db_xref="GI:3516"
/db_xref="SWISS-PROT:P06100"
/translation="MEKFGKALVPLLLKLEDEKLSSTYDHSMTLGCADLSSMLYSLGIP
RDSQHRVLDITFQSWAETSSEVPEFTFESFTNIPGVLOSTVTPPCFNISQNDQ
RVALFQDETLLFLFKHPTVIQELTYLELRKRNRYHKTILKAWLTQDPMPEIVSAD
GLSERGSYVFFDPQWRKQRCQDFLLFYNAIM"
3135..3137
/gene="CDC33"
/note="minor 3' polyadenylation site"
/db_xref="SGD:S0002324"
3142..3144
/gene="CDC36"
/note="major 3' polyadenylation site"
/db_xref="SGD:S0002324"
771 g 1158 t
BASE COUNT 1165 a 755 c 771 g 1158 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 7; Length 3849;
Best Local Similarity 89.5%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACATATGAGCCTTCATG 20
||||||| | | | | | | |
Db 3611 TACATATGACCATTCATG 3593

RESULT 14
SCYDL165W 944 bp DNA PLN 11-AUG-1997
LOCUS S.cerevisiae chromosome IV reading frame ORF YDL165W.
DEFINITION 274213..271256
ACCESSION 274213..271256
NID 91431261
VERSION 274213.1 GI:1431261
KEYWORDS
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
REFERENCE 1 (bases 1 to 944)
AUTHORS Pohl,T.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 944)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-1996) Data collected by MIPS on behalf of the

European yeast chromosome IV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org
Location/Qualifiers
1..944
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="IV"
248..823
/gene="CDC36"
/db_xref="SGD:S0002324"
248..823
/gene="CDC36"
/note="ORF YDL165W"
/codon_start=1
/db_xref="SGD:S0002324"
/protein_id="CAA98739.1"
/db_xref="PID:e253074"
/db_xref="PID:g1431262"
/db_xref="GI:1431262"
/db_xref="SWISS-PROT:P06100"
/translation="MEKFGKALVPLLLKLEDEKLSSTYDHSMTLGCADLSSMLYSLGIP
RDSQHRVLDITFQSWAETSSEVPEFTFESFTNIPGVLOSTVTPPCFNISQNDQ
RVALFQDETLLFLFKHPTVIQELTYLELRKRNRYHKTILKAWLTQDPMPEIVSAD
GLSERGSYVFFDPQWRKQRCQDFLLFYNAIM"
315 a 171 c 178 g 280 t
BASE COUNT 315 a 171 c 178 g 280 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 7; Length 944;
Best Local Similarity 89.5%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACATATGAGCCTTCATG 20
||||||| | | | | | | |
Db 313 TACATATGACCATTCATG 331

RESULT 15
SCSFAARP 7008 bp DNA PLN 29-NOV-1994
LOCUS S.cerevisiae SFA and ARP genes.
DEFINITION X68020 S59849
ACCESSION X68020 S59849
NID g577609
VERSION X68020.1 GI:577609
KEYWORDS arp gene; beta-D-galactosidase; CDC36 gene; CDC9 gene; sfa gene.
SOURCE baker's yeast.
ORGANISM Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
REFERENCE 1 (bases 1 to 7008)
AUTHORS Wehner,E.P., Rao,E. and Brendel,M.
TITLE Molecular structure and genetic regulation of SFA, a gene
responsible for resistance to formaldehyde in Saccharomyces
cerevisiae, and characterization of its protein product
Mol. Gen. Genet. 237 (3), 351-358 (1993)
JOURNAL 93247548
REFERENCE 2 (bases 1 to 7008)
AUTHORS Wehner,E.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-1992) E. Wehner, Institut f Mikrobiologie,
Theodor Steinkai 7, Haus 75, 6000 Frankfurt/Main, FRG
REFERENCE 3 (bases 1 to 7008)
AUTHORS Wehner,E.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-1994) E. Wehner, Institut f Mikrobiologie,
Theodor Steinkai 7, Haus 75, 6000 Frankfurt/Main, FRG
COMMENT On Nov 28, 1994 this sequence version replaced gi:288588.
FEATURES
Location/Qualifiers
1..7008
/organism="Saccharomyces cerevisiae"
/strain="AH22"
/db_xref="taxon:4932"
```

/chromosome="IV"

gene

1. .442

CDS

/gene="CDC9"

CDS

<1. .442

CDS

/gene="CDC9"

CDS

/codon_start=2

CDS

/protein_id="CAA48158.1"

CDS

/db_xref="PID:g600039"

CDS

/db_xref="GI:600039"

CDS

/db_xref="SWISS-PROT:P04819"

CDS

/translation="DLGVGAYGRKRTGTGGFLGVCYNQDTEGETCKIGTGFS

CDS

DEMLQLHLDRTPTIIDPKATVFDSSAPDVNFETTLFEVLTAADLSLSPYKAGS

CDS

ATFDKGVSLREPIFLRIREDKGVEDATSSQIVELYENQSHMQN"

CDS

complement(565..1140)

CDS

/gene="CDC36"

CDS

/db_xref="SGD:S0002324"

CDS

/db_xref="PID:g577610"

CDS

/gene="CDC36"

CDS

/codon_start=1

CDS

/db_xref="SGD:S0002324"

CDS

/protein_id="CAA48160.1"

CDS

/db_xref="PID:g577610"

CDS

/db_xref="GI:577610"

CDS

/db_xref="SWISS-PROT:P06100"

CDS

/translation="MEKFGKALVPLLLKLEKLSSTYDHSMTLGADLSMMLYSLGIP

CDS

RDSQDHYLDTFQSPWAEFSRSEVEPRFFTFPIPGVLOSTVTPPCFNSIONDOQ

CDS

RYALQDTEFLFYKHPGTVIQELTYLELRKNRYHKILKAWLTKDPMMEPIVSAD

CDS

GLSEKSYFFDPQWKEKQKQDFLLFYNAIM"

CDS

2277. .4436

CDS

/gene="ARP"

CDS

/db_xref="SGD:S0002326"

CDS

2277. .4436

CDS

/gene="ARP"

CDS

/codon_start=1

CDS

/db_xref="SGD:S0002326"

CDS

/protein_id="CAA48159.1"

CDS

/db_xref="PID:g288590"

CDS

/db_xref="GI:288590"

CDS

/db_xref="SWISS-PROT:P32770"

CDS

/translation="MHYVLELOVAHLDPDKDOCRITANIAFOIVNAETLVCHYGTNS

CDS

LPSIYVNGTITLSLSAMVQLDKIDHVIQNDDFVLVSLYSTWHIRVTLPRQARDGFI

CDS

LTSYLOHPKFDLWDFKVCNVHPEILQKKAISNNNCNTKSIISAAKNTKDLDEI

CDS

VRILEYPTSEAGSVPEIYSLKRTDILQLHKKCTSPEDMESVLTQKPYDSHTDIR

CDS

AFLOEKSLILYNNLPDPTQSELESWFQYGVPRGVFVKNIVEDETSNNVNNNSLN

CDS

NSPYVEDDQSLSGFVFOVTHEEATVIALNGRSILSLNANTKQPRVVEHVELOPSSST

CDS

GVLDKAEILSPFPQSKNKPREGDNCPCSGFSNFORCTACFCSPAPNSQIHTAN

CDS

SNVNNYSRNLNRYNSGSSNISNTAANRPYCAPEFNMIANTPAALTYNRAHFFA

CDS

ITPLSRQSLNMAPSNSGSPILIIADHFSGNNNIAPNRYNNNNNNNNNNNNNNNNRY

CDS

NINNINGNGNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGNNGN

CDS

SNIGMGCGSNMFRAGDKWCTCTYHFNKAVYVLCRCGGPKISGDSATNHYIDS

CDS

TFGPASRTSPNNISVNTNGSNAGRTDGNKGRDISLMEFMSPLSMATKSMKEGD

CDS

GNSSSFNEFKSKANVNFNVYGDNSAFNGFNSSIRW"

CDS

complement(4667..5827)

CDS

/gene="SFA"

CDS

/db_xref="SGD:S0002327"

CDS

/db_xref="PID:g288591"

CDS

/gene="SFA"

CDS

/codon_start=1

CDS

/db_xref="SGD:S0002327"

CDS

/protein_id="CAA48161.1"

CDS

/db_xref="GI:288591"

CDS

/db_xref="SWISS-PROT:P32771"

CDS

/translation="MSAATVGKPIKICAAVAYDAKKPLSVEEITVDAPKAHEVRKIE

CDS

YVAVCHTAYTLSGSDPEGLPCVLGHEGAGIVSGVDVITVKPGDHVIALYTAECG

CDS

KCKCTSGKTLGAVRATQGVMPDGTREHNAKGEDYHEMGCSTFSEYTVVADV

CDS

SVVADPRAPLDAACLLCGVTTFGAALKATYQKGDYAVFGCGTVGLSVIQAKL

CDS

RGASLIADINNNKKQYCSQFATDFVNPREDLAKDQITVEKLIEMTDGGLDFDC

CDS

TGNTIMRDALEACHKGWGQIIIGVAAAGEEISTRPQLVTGRWKGSAFGGKGRS

CDS

EMGGLIKDYQKALKEEFITRRPFEINQAFEDLHNGDCLRTVLKSDIEK"

CDS

2266 a. 1305 c 1367 g 2070 t

Query Match

Best Local Similarity

Matches 17; Conservative

0; Mismatches

2; Indels

0; Gaps

0;

QY 2 TACATATGAGCCTTCCATG 20

Db 1075 TACATATGACCATTCATG 1057

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

|||||

Score 15.8; DB 8; Length 7008;

Pred. No. 79;

Mismatches 0;

Indels 2;

Gaps 0;

Search completed: September 18, 1999, 15:49:46

Job time: 1466 sec

GenCore version 4.5
Copyright (C) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 05:27:26 : Search time 213.04 Seconds
(without alignments)
23.488 Million cell updates/sec

Title: US-09-037-472-10
Perfect score: 20
Sequence: 1 TTACATATGAGCCTTCCATG 20

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	1	TT13884
2	20	100.0	20	1	V32396
3	20	100.0	20	1	Interleukin-1-alpha
4	18.4	92.0	5701	1	Human IL-1ra BAC c
5	16.4	82.0	6285	1	Urea amidolase ge
6	15.8	79.0	4708	1	Human acetyl choli
7	15.8	79.0	7168	1	HEV strain SAR-55
8	15.8	79.0	7168	1	Hepatitis E virus
9	15.8	79.0	7168	1	Hepatitis E virus
10	15.2	76.0	354	1	Expressed Sequence
11	15.2	76.0	354	1	Human brain Expre
12	15.2	76.0	8148	1	Streptococcus pneu
13	15.2	76.0	36	1	PCR primer used to
14	14.8	74.0	1891	1	Potato citrate syn
15	14.8	74.0	1747	1	Tobacco citrate sy
16	14.8	74.0	1891	1	Potato citrate syn
17	14.8	74.0	110000	1	Continuation (2 of
18	14.8	74.0	110000	1	V30458_1
19	14.2	71.0	7833	1	Human papilloma vi
20	14.2	71.0	2387	1	Brh-I cDNA, Polype
21	14.2	71.0	8841	1	Tryptophan-2,3-dio
22	14.2	71.0	4782	1	Rat homologue of h
23	14.2	71.0	33	1	Human immunodefici
24	14.2	71.0	2387	1	Wasp Brh-1 toxin g
25	14.2	71.0	467	1	Human papillomavir
26	14.2	71.0	288	1	Pyrococcus furiosu
27	14.2	71.0	467	1	DNA sequence of LI
28	14.2	71.0	2846	1	Human TRAF5 cDNA
29	14.2	71.0	3993	1	Human papilloma vi
30	14.2	71.0	455	1	Rat kidney cell ca
31	14.2	71.0	4131	1	HIV-1 proviral DNA
32	14.2	71.0	33	1	V20488
33	14.2	71.0	4131	1	Rat kidney calcium
34	14.2	71.0	145	1	Target sequence in
35	14.2	71.0	1570	1	Ehrlichia canis VS
36	14.2	71.0	3168	1	Candida neoformans
37	14.2	71.0	3220	1	Candida neoformans
38	14.2	71.0	3168	1	Candida neoformans
39	14.2	71.0	610	1	Staphylococcus aur
40	14.2	71.0	4131	1	Rat parathyroid ca
41	14.2	71.0	4718	1	Mouse semaphorin r
42	14.2	71.0	4733	1	Mouse semaphorin r
43	14.2	71.0	4769	1	Mouse semaphorin r

C 44 14.2 71.0 4784 1 X09000 Mouse semaphorin r
C 45 14.2 71.0 864 1 X34760 DNA encoding p30a

ALIGNMENTS

RESULT 1

TT13884
ID TT13884 standard; DNA; 20 BP.
AC TT13884;
DT 19-OCT-1997 (first entry)
DE Primer for detecting genetic predisposition to periodontal disease.
KW Periodontal disease; gingivitis; periodontitis; polymorphism;
KW Interleukin-1 alpha; IL-1A; interleukin-1 beta; IL-1B; primer; PCR;
KW Polymerase chain reaction; ss.
OS Synthetic.
PN WC9706180-A1.
PD 20-FEB-1997.
PF 02-AUG-1996; U12455.
PR 03-AUG-1995; US-510696.
PA (KORN/) KORNMAN R S.
PA (MEDI-) MEDICAL SCI SYSTEMS INC.
PI Duff GW, Kornman KS;
DR WPI; 97-154207/14.
PT Identification of patient's genetic polymorphism pattern - allows
PT prediction of increased periodontal disease severity
PS Claim 3; Page 25; 33pp; English.
CC PCR primers (T70316 and TT13884-88) can be used in a new method
CC for identifying a genetic predisposition to periodontal disease by
CC detecting the presence of DNA polymorphisms in the gene sequences
CC for interleukin-1 alpha (IL-1A) and interleukin-1 beta (IL-1B).
CC Alleles associated with severe disease were identified as IL-1A
CC allele 2 together with IL-1B (raql) allele 2. The primers are used
CC to amplify DNA from a blood or tissue sample and products are
CC subjected to restriction digestion to determine the polymorphism
CC pattern of the patient. A single base variation (C/T) polymorphism
CC at IL-1A base -889 can be identified using primers (T70316 and
CC TT13884) corresponding to bases -967 to -945 and -888 to -869,
CC respectively. An NcoI site is created if C is available at -889,
CC but not if T is present.
SQ Sequence 20 BP; 5 A; 5 C; 3 G; 7 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.062; 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0; Indels 0;

QY 1 TTACATATGAGCCTTCCATG 20

|||||
Db 1 TTACATATGAGCCTTCCATG 20

RESULT 2

V32396
ID V32396 standard; DNA; 20 BP.
AC V32396;
DT 11-SEP-1998 (first entry)
DE Interleukin-1-alpha primer 4 (-888/-869).
KW IL-1-alpha; genetic polymorphism; PCR; primer; amplification;
KW slight threatening diabetic retinopathy; interleukin-1-alpha;
KW Interleukin-1-beta; interleukin-1RN; ss.
OS Synthetic.
OS Homo sapiens.
PN WC9815653-A1.
PD 16-APR-1998.
PF 09-OCT-1997; G02790.
PR 10-OCT-1996; GB-021129.
PA (DUFF/) DUFF G.
PA (RENN/) RENNIE I.
PA (RICH/) RICHARDSON R.
PI Duff G, Rennie I, Richardson R;
DR WPI; 98-240835/21.

PT Predicting increased risk of sight-threatening diabetic retinopathy
 PT - comprises identifying genetic polymorphism pattern for genes
 PT IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
 PT symptoms occur
 PS Claim 2: Page 33; 41pp; English.
 CC Interleukin-1-alpha (IL-1-alpha) primers 4 and 3 (V32395) were used to
 CC amplify the IL-1-alpha gene region to identify single base variation
 CC polymorphism of C/T at base 889. The invention claims to provide a
 CC method for predicting the risk of sight threatening diabetic retinopathy.
 CC The method involves isolating DNA from a patient and determining the DNA
 CC polymorphism pattern of the genes that code for interleukin-1-alpha,
 CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
 CC identified is then compared with controls of known DNA polymorphism
 CC patterns thereby identifying patients carrying a genetic polymorphism
 CC associated with increased risk of sight threatening diabetic retinopathy.
 CC The method may be able to identify diabetic patients at risk before the
 CC clinically detectable disorders occur. Polymorphism pattern
 CC determination of IL genes involved PCR reactions using primers V32389-
 CC V32398. The method is also claimed to be useful in conjunction with
 CC identification of other genes associated with sight threatening diabetic
 CC retinopathy in genomic DNA and therefore, in identifying diabetic
 CC patients expressing multiple risk patterns.
 SQ Sequence 20 BP; 5 A; 5 C; 3 G; 7 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.062;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCCATG 20
 |||||
 Db 1 TTACATATGAGCCTTCCCATG 20

RESULT 3
 X16612

ID X16612 standard; DNA; 20 BP.
 AC X16612;
 DT 29-APR-1999 (first entry)
 DE Interleukin 1 (44112332) haplotype PCR primer #6.
 KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;
 KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;
 KW Graves disease; systemic lupus erythematosus; lichen sclerosis;
 KW ulcerative colitis; PCR primer; ss.
 OS Synthetic.
 OS Homo sapiens.
 PN W09854359-A1.
 PD 03-DEC-1998.
 PF 21-MAY-1998; G01481.
 PR 29-MAY-1997; GB-011040.
 PA (CAMP/) CAMP N J.
 PA (COXA/) COX A.
 PA (DGIO/) DE GIOVINE F S.
 PA (DUFF/) DUFF G.
 PI Camp ND, Cox A, De Giovine FS, Duff G;
 DR WPI; 99-080814/07.
 PT New method of determining a patient's susceptibility to inflammatory
 PT disorders - by detecting the presence of an IL-1 (44112332)
 PT haplotype, useful in designing treatment strategies that modulate
 PT the activity of proteins produced by the IL-1 gene cluster
 PS Claim 3; Page 33; 49pp; English.
 CC A method has been developed for determining a patient's susceptibility
 CC to an inflammatory disorder. The method comprises the detection of an
 CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the
 CC patient, where its presence indicates susceptibility to an inflammatory
 CC disorder. X16607 to X16631 represent PCR primer used in the method for
 CC detecting the IL-1 (44112332) haplotype. The method provides kits for
 CC the early prediction of a patient's susceptibility to inflammatory
 CC disorders, including coronary artery disease, osteoporosis, nephropathy
 CC in diabetes mellitus, alopecia areata, Graves disease, systemic lupus
 CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of
 CC alleles of the haplotype can be applied to particular inflammatory
 CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,

CC psoriasis, and insulin dependent diabetes. The identification of a
 CC disease-associated haplotype enables determination of which alleles are
 CC causative, and this information is useful in designing treatment
 CC strategies, including gene therapy and treatment using various agents
 CC that modulate the activity of proteins produced by the IL-1 gene cluster.
 CC Some alleles from the IL-1 gene cluster are associated with particular
 CC inflammatory diseases, and insufficient IL-1 production appears to act
 CC centrally in the pathology of these diseases. Therefore, the use of IL-1
 CC gene clusters is useful in determining genetic susceptibility to
 CC inflammatory diseases, including those with a multifactorial etiology
 CC with a polygenic component.
 SQ Sequence 20 BP; 5 A; 5 C; 3 G; 7 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.062;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCCATG 20
 |||||
 Db 1 TTACATATGAGCCTTCCCATG 20

RESULT 4

X02988/C
 ID X02988 standard; DNA; 5701 BP.
 AC X02988;
 DT 22-JUN-1999 (first entry)
 DE Human IL-1ra BAC contiguous DNA sequence 33.
 KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
 KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
 KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
 KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
 OS Homo sapiens.
 PN W09906426-A1.
 PD 11-FEB-1999.
 PF 03-AUG-1998; I16102.
 PR 02-JUL-1998; US-091650.
 PR 04-AUG-1997; US-054646.
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 PI Pan Y;
 DR WPI; 99-153692/13.
 PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
 PT used to inhibit inflammation and to screen for specific modulators
 PT Example 5; Figure 3; 226pp; English.
 PS X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
 CC containing alternatively spliced forms of human IL-1ra. Such fragments
 CC are used in the method of the invention which describes the isolation of
 CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
 CC member of the cytokine superfamily that is expected to inhibit
 CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
 CC also bind to a new receptor so could regulate other cellular processes
 CC associated with acute or chronic inflammation, e.g. asthma, chronic
 CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
 CC bowel disease. It may also induce or suppress interleukins, cytokines and
 CC growth factors. Modulators of this protein are used to treat or prevent
 CC conditions associated with abnormal levels of inflammation, or activity
 CC of IL-1 or its receptor complex.
 SQ Sequence 5701 BP; 1729 A; 1238 C; 1079 G; 1654 T;

Query Match 92.0%; Score 18.4; DB 1; Length 5701;
 Best Local Similarity 95.0%; Pred. No. 0.81;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCCATG 20
 |||||
 Db 529 TTACATATGAGCCTTCAATG 510

RESULT 5

Q49460
 ID Q49460 standard; cDNA; 6265 BP.

AC Q49460;
 DT 10-MAY-1994 (first entry)
 DE Urea amidolyase gene.
 KW Urea amidolyase; URL; yeast; recombinant plasmid; ss.
 OS Saccharomyces cerevisiae.
 FH Key
 FT cds
 FT Location/Qualifiers
 FT 627..6134
 FT /tag= a
 FT /product= URL
 PN J05244959-A.
 PD 24-SEP-1993.
 PF 05-MAR-1992; 084531.
 PR 05-MAR-1992; JP-084531.
 PA (TOYM) TOYOBO KK.
 DR WPI: 93-338925/43.
 P-PSDB: R42839.
 PT DNA having the genetic information of urea amidolyase originated from Saccharomyces yeast - can be used to prepare high purity urea amidolyase by culturing the transformant comprising the DNA Claim 1; Page 10-17; 17pp; Japanese.
 PS This sequence encodes a protein which has urea amidolyase (URL)-CC activity and is derived from yeast. This sequence may be used within a recombinant plasmid for the production of highly pure URL.
 CC a recombinant plasmid for the production of highly pure URL.
 SQ Sequence 6265 BP; 1873 A; 1156 C; 1361 G; 1875 T;

Query Match 82.0%; Score 16.4; DB 1; Length 6265;
 Best Local Similarity 94.4%; Pred. No. 9.2;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCA 18
 || |||||
 Db 937 TTGACATATGAGCCTTCCA 954

RESULT 6
 NS0415/c
 ID NS0415 standard; DNA; 4708 BP.
 AC NS0415;
 DT 08-JAN-1992 (first entry)
 DE Human acetyl choline receptor alpha gene.
 KW hACR-alpha; myasthenia gravis; MG; neuromyopathy; ds.
 OS Homo sapiens.
 FH Key
 FT cds
 FT Location/Qualifiers
 FT 46..3126
 FT /tag= a
 FT mat_peptide 346..3123
 FT /tag= b
 FT intron 39..328
 FT /tag= c
 FT /note= "60N inserted as a spacer"
 FT intron 475..585
 FT /tag= d
 FT intron 531..917
 FT /tag= e
 FT /note= "60N inserted as spacer"
 FT intron 1028..1200
 FT /tag= f
 FT /note= "60N inserted as spacer"
 FT intron 1397..1624
 FT /tag= g
 FT /note= "60N inserted as spacer"
 FT intron 1863..2051
 FT /tag= h
 FT /note= "60N inserted as spacer"
 FT intron 2276..2430
 FT /tag= i
 FT /note= "60N inserted as spacer"
 FT intron 2671..2994
 FT /tag= j
 PN J60078996-A.
 PD 04-MAY-1985.
 PF 05-OCT-1983; 186402.

PR 05-OCT-1983; JP-186402.
 PA (MITU) MITSUBISHI CHEM IND KK.
 DR WPI: 85-144120/24.
 P-PSDB: P50361.
 FT New peptide(s) useful in treatment of myasthenia gravis - also in diagnosis, having acetylcholine receptor-alpha like activities.
 PS Disclosure; Fig 1; 13pp; Japanese.
 CC Peptides derived from the hACR-alpha gene product have similar immunogenicity to the gene product, and may be useful in the treatment and diagnosis of myasthenia gravis caused by neuropathy in autoimmune reaction to ACR.
 CC The sequence given contains breaks in some of the introns, at these locations 60 N residues have been inserted to preserve integrity.
 SQ Sequence 4708 BP; 1124 A; 1059 C; 965 G; 1200 T;

Query Match 79.0%; Score 15.8; DB 1; Length 4708;
 Best Local Similarity 89.5%; Pred. No. 19;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCAT 19
 || |||||
 Db 1164 TGACAGATGAGCCTTCCAT 1146

RESULT 7
 Q45197
 ID Q45197 standard; cDNA; 7168 BP.
 AC Q45197;
 DT 21-OCT-1994 (first entry)
 DE HEV strain SAR-55 cDNA sequence.
 KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF; antibody; detection; diagnosis; primates; stool suspension; ss.
 OS Hepatitis E virus strain SAR-55
 FH Key
 FT Location/Qualifiers
 FT 28..5109
 FT /tag= a
 FT misc_difference 3739..3741
 FT /tag= b
 FT /codon= seq:cag, aa:Glu
 FT misc_difference 3757..3759
 FT /tag= c
 FT /codon= seq:cag, aa:Glu
 FT misc_difference 4081..4083
 FT /tag= d
 FT /codon= seq:gtg, aa:Glu
 FT misc_difference 5011..5013
 FT /tag= e
 FT /codon= seq:ggc, aa:Glu
 FT cds 5147..7129
 FT /tag= f
 FT /label= ORF-2
 FT misc_difference 5780..5782
 FT /tag= g
 FT /codon= seq:tgg, aa:Tyr
 FT cds 5106..5457
 FT /tag= h
 FT /label= ORF-3
 PN W09406913-A.
 PD 31-MAR-1994.
 PF 17-SEP-1993; U08849.
 PR 18-SEP-1992; US-947263.
 PA (USSH) US SEC DEPT HEALTH.
 PI Emerson SU, Purcell RH, Tsarev SA;
 DR WPI: 94-118462/14.
 P-PSDB: R51284-66.
 PT Purified Hepatitis E strain SAR-55 virus - used to develop prods. for use in detection, diagnosis, vaccines and therapy of hepatitis E virus infection
 PS Claim 2; Page 16-20; 114pp; English.
 CC This sequence represents the genomic sequence of the hepatitis E virus (HEV) strain SAR-55. This sequence contains three open reading

CC frames (ORFs). The proteins encoded by this sequence can be used to
CC stimulate the production of protective antibodies upon injection into
CC a mammal that would serve to protect the mammal upon challenge with
CC wild type HEV. The proteins can be used for detection and diagnosis
CC of HEV infection. This cDNA was isolated from primates inoculated
CC with stool suspensions obtained from hepatitis E patients.
SQ Sequence 7168 BP; 1223 A; 2294 C; 1867 G; 1784 T;

Query Match 79.0%; Score 15.8; DB 1; Length 7168;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCAT 19
||||| ||||||||| ||
Db 2189 TTACATCTGAGCCTTCTAT 2207

RESULT 8
T27394
ID T27394 standard; cDNA; 7158 BP.
AC T27394;
DT 26-NOV-1996 (first entry)
DE Hepatitis E virus strain SAR-55 cDNA (ATCC 75302).
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW structural region; antigen; detection; antibody; vaccine;
KW immunisation; infection; ss.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT cds 28..5099
FT /*tag= a
FT /label= ORF-1 (R91813)
FT /transl_except= pos:3739..3741, aa:Glu
FT /note= "10 bp nucleic acid sequence TGGTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 for numbering to conform to
FT that given in the specification".
FT cds 5137..7119
FT /*tag= b
FT /label= ORF-2 (R91814)
FT /note= "10 bp nucleic acid sequence TGGTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 for numbering to conform to
FT that given in the specification".
FT cds 5096..5467
FT /*tag= c
FT /label= ORF-3 (R91815)
FT /note= "10 bp nucleic acid sequence TGGTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 for numbering to conform to
FT that given in the specification".
PN WO9610580-A2.
PD 11-APR-1996.
PF 03-OCT-1995; U13102.
PR 03-OCT-1994; US-316765.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 96-209320/21
DR P-PSDB; R91813, R91814, R91815.
PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT antigenic protein useful in diagnosis, prophylaxis and treatment of
PT hepatitis E virus infection
PS Claim 2: Pages 16-21; 121pp; English.
CC The present sequence is the cDNA of the hepatitis E virus (HEV)
CC strain SAR-55, which was implicated in an enterically transmitted
CC non-A, non-B hepatitis in Pakistan. The protein encoded by the
CC structural region of the virus (i.e. ORF-2), which is capable of
CC forming HEV like particles, is useful for the detection of HEV
CC antibodies (pref. IgG or IgM) in blood, plasma, sera,
CC cerebrospinal fluid, tissue, urine or pleural fluid. The protein,
CC and anti-HEV antibodies generated using the protein, can also be
CC used in vaccines for immunising an animal against HEV infection.
CC The protein is identified as a band of greater than 50 kD

CC following SDS-PAGE of cell lysates of insect cells infected with
CC a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
CC expression vectors pPIC9-1779, -1780 and -1781.
SQ Sequence 7158 BP; 1221 A; 2293 C; 1864 G; 1780 T;

Query Match 79.0%; Score 15.8; DB 1; Length 7158;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCAT 19
||||| ||||||||| ||
Db 2189 TTACATCTGAGCCTTCTAT 2207

RESULT 9
V71604
ID V71604 standard; DNA; 7168 BP.
AC V71604;
DT 02-FEB-1999 (first entry)
DE Hepatitis E virus (HEV) polypeptides encoding nucleic acid SAR-55.
KW Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW passive immunisation; ss.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT cds 28..5109
FT /*tag= a
FT /transl_except= (pos:3739..3741, aa:Glu)
FT /transl_except= (pos:3757..3759, aa:Glu)
FT /transl_except= (pos:4081..4083, aa:Glu)
FT /transl_except= (pos:5011..5013, aa:Glu)
FT /product= "ORF-1 protein"
FT 5147..7129
FT /*tag= b
FT /transl_except= (pos:5780..5782, aa:Tyr)
FT /product= "ORF-2 protein"
FT 5106..5477
FT /*tag= c
FT /product= "ORF-3 protein"
PN WO9846761-A1.
PD 22-OCT-1998.
PF 09-APR-1998; U07418.
PR 11-APR-1997; US-840316.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Robinson RA, Tsarev SA;
DR WPI: 98-568733/48.
DR P-PSDB; W81519, W81520, W81521.
PT New hepatitis E virus DNA from Pakistani strain SAR-55 - used for,
PT e.g. developing products for diagnosis of, and vaccination against
PT hepatitis E virus infection
PS Disclosure: Pages 126-131; 204pp; English.
CC This represents a DNA sequence designated SAR-55 encoding hepatitis E
CC virus (HEV) ORF proteins ORF-1, ORF-2 and ORF-3. A host organism
CC transformed or transfected with a recombinant expression vector
CC containing the SAR-55 nucleic acid can be used to produce the HEV
CC proteins, especially ORF-2 protein. The recombinant HEV proteins can
CC used as diagnostic agents and as vaccines for use against HEV infection.
CC The detection of antibodies specific for HEV can be used for the
CC diagnosis of infection and diseases caused by HEV, and for monitoring the
CC progression of such disease. Such methods are also useful for monitoring
CC the efficacy of therapeutic agents during the course of treatment of HEV
CC infection and disease in a mammal. The antibodies can be used for
CC detection or for passive immunisation of mammals.
SQ Sequence 7168 BP; 1222 A; 2294 C; 1868 G; 1784 T;

Query Match 79.0%; Score 15.8; DB 1; Length 7168;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCAT 19
||||| ||||||||| ||
Db 2189 TTACATCTGAGCCTTCTAT 2207

DT 22-OCT-1998 (first entry)
 DE PCR primer used to mutate part of genome of adenovirus serotype 5.
 KW Domain alpha 2; antigen; major histocompatibility complex class I;
 KW MHC-I; primary receptor; serotype C adenovirus; human; fibronectin;
 KW module III; targeting; cell-surface protein; cellular receptor;
 KW Bifunctional ligand; tumour cell; infected cell; PCR primer; ss.
 OS Synthetic.
 OS Mastadenovirus.
 PN FR2758822-AL.
 PD 31-JUL-1998.
 PF 09-SEP-1997; 011166.
 PR 30-JAN-1997; FR-001005.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PI Boulanger P, Hong SS, Karayan L;
 DR WPI; 98-416459/36.
 PT Targeting adenovirus to cell-surface protein - using bifunctional
 PT ligand that interacts with adenovirus fibre and cell-surface protein
 PS Example 1; Page 34; 43pp; French.
 CC PCR primers V42902-03 were used to amplify part of the genome of
 CC adenovirus serotype 5, and introduce NcoI and KpnI sites for cloning. It
 CC is believed that domain alpha 2 major histocompatibility complex class I
 CC (MHC-I) constitutes the primary receptor for serotype C adenoviruses,
 CC and that human fibronectin module III (sic) functions as a co-receptor
 CC or cofactor. The specification describes a method for targeting
 CC adenovirus to cell-surface proteins other than the natural cellular
 CC receptor of the adenovirus using a bifunctional ligand. The bifunctional
 CC ligand comprises a portion capable of interacting with the adenovirus,
 CC a portion capable of interacting with the cell-surface protein, and
 CC optionally a spacer between the first and second portions. The ligand is
 CC used to target an adenovirus to a host cell of interest, e.g. a tumour
 CC cell, an infected cell, or a particular cell type bearing a specific
 CC surface marker.
 SQ Sequence 36 BP; 6 A; 9 C; 7 G; 14 T;

Query Match 76.0%; Score 15.2; DB 1; Length 36;
 Best Local Similarity 85.0%; Pred. No. 22;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCATG 20
 ||||| ||| |||||
 DB 5 TTACATTTGACTCTTCATG 24

RESULT 14
 T04199
 ID T04199 standard; cDNA to mRNA; 1891 BP.
 AC T04199;
 DT 25-JAN-1996 (first entry)
 DE Potato citrate synthase cDNA.
 KW Citrate synthase; flower formation; tuber storage; ss.
 OS Solanum tuberosum.
 FH Key Location/Qualifiers
 FT cds 73..1485
 FT /*tag= a
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR P-PSDB; R82838.

PT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Claim 5; Page 53-56; 87pp; English.
 CC To identify a cDNA from potato which codes for citrate synthase, a
 CC cDNA fragment of citrate synthase from Arabidopsis thaliana was
 CC firstly amplified using A.thaliana cDNA and oligos T04202 and T04203
 CC which are complementary to the 5' or 3' end of the coding region of

CC A. thaliana cDNA for citrate synthase. The oligos additionally
 CC introduce BamHI cleavage sites at both ends of the amplified cDNA
 CC fragment. A cDNA library was prepd. from potato leaves and screened
 CC with A. thaliana citrate synthase cDNA. Positive clones were
 CC purified and sequenced. The nt sequence is given in T04199.
 SQ Sequence 1891 BP; 512 A; 370 C; 425 G; 584 T;

Query Match 74.0%; Score 14.8; DB 1; Length 1891;
 Best Local Similarity 88.9%; Pred. No. 56;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ACATATGAGCCTTCATG 20
 ||||| ||| |||||
 DB 688 ACATATGAGGATTCATG 705

RESULT 15

T04201
 ID T04201 standard; cDNA to mRNA; 1747 BP.
 AC T04201;
 DT 25-JAN-1996 (first entry)
 DE Tobacco citrate synthase cDNA.
 KW Citrate synthase; flower formation; ss.
 OS Nicotiana tabacum
 FH Key Location/Qualifiers
 FT cds 70..1476
 FT /*tag= a
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschutze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR P-PSDB; R82840.
 PT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Claim 7; Page 60-63; 87pp; English.
 CC To identify a cDNA from tobacco which codes for citrate
 CC synthase, a cDNA bank of leaf tissue from tobacco was prepd.
 CC plaques of this cDNA bank were screened using a radioactive DNA
 CC probe which comprises Solanum tuberosum citrate synthase cDNA
 CC (T04199). One of the clones was sequenced. The nt. sequence is
 CC given in T04201.
 SQ Sequence 1747 BP; 490 A; 335 C; 400 G; 522 T;

Query Match 74.0%; Score 14.8; DB 1; Length 1747;
 Best Local Similarity 88.9%; Pred. No. 56;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ACATATGAGCCTTCATG 20
 ||||| ||| |||||
 DB 682 ACATATGAGGATTCATG 699

Search completed: September 18, 1999, 05:27:27
 Job time: 1692 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:49 ; Search time 1405 Seconds
(without alignments)
28.079 Million cell updates/sec

Title: US-09-037-472-10
Perfect score: 20
Sequence: 1 TTACATATGAGCCTTCCATG 20
Scoring table: IDENTITY_NUC
Searched: 2546578 seqs, 986266752 residues

Database :
1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match %	Query Length	DB ID	Description
C 1	17.4	87.0	407	21	R19433
C 2	16.8	84.0	360	23	D55860
C 3	16.8	84.0	476	24	N29325
C 4	16.8	84.0	488	39	AA831249
C 5	16.8	84.0	493	40	AA908884
C 6	16.8	84.0	598	42	AA130847
C 7	16.8	84.0	567	44	A1248140
C 8	16.4	82.0	379	23	D52258
C 9	16.4	82.0	402	23	D53341
C 10	15.8	79.0	447	24	H90734
C 11	15.8	79.0	424	24	H95824
C 12	15.8	79.0	630	30	AA197351
C 13	15.8	79.0	427	31	AA275311
C 14	15.8	79.0	432	31	AA288092
C 15	15.8	79.0	372	32	AA377278
C 16	15.8	79.0	659	37	AA673147
C 17	15.8	79.0	418	38	AA748850
C 18	15.8	79.0	525	38	AA790389
C 19	15.8	79.0	485	39	AA863883
C 20	15.8	79.0	475	43	A1226166
C 21	15.4	77.0	405	23	R93797
C 22	15.4	77.0	474	24	H78857
C 23	15.4	77.0	396	33	AA401166
C 24	15.2	76.0	354	20	M61970
C 25	15.2	76.0	400	21	D46229
C 26	15.2	76.0	490	22	H08527
C 27	15.2	76.0	438	22	H14725
C 28	15.2	76.0	464	22	R51062
C 29	15.2	76.0	275	22	R66347
C 30	15.2	76.0	301	22	R67490
C 31	15.2	76.0	450	24	H73109
C 32	15.2	76.0	447	24	H82720
C 33	15.2	76.0	326	24	N36040
C 34	15.2	76.0	241	25	N46516
C 35	15.2	76.0	408	25	N62510
C 36	15.2	76.0	446	27	AA028135
C 37	15.2	76.0	416	29	AA126072
C 38	15.2	76.0	501	29	AA194540
C 39	15.2	76.0	414	30	AA270555
C 40	15.2	76.0	320	31	AA280522
C 41	15.2	76.0	474	31	AA290446
C 42	15.2	76.0	256	32	AA379224
C 43	15.2	76.0	371	33	AA434013
C 44	15.2	76.0	317	33	AA442950
C 45	15.2	76.0	461	51	A0069310

ALIGNMENTS

RESULT 1
R19433/c
LOCUS
Yg25c03.r1 Soares infant brain 1N1B Homo sapiens cDNA clone
IMAGE:33575 5', mRNA sequence.
R19433
9773043
R19433.1 GI:773043
EST 14-APR-1995

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 407)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 2026
High quality sequence stops: 228 Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 2026 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 228.
Location/Qualifiers
1..407
/organism="Homo sapiens"
/db_xref="GDB:405922"
/db_xref="taxon:9606"
/clone="IMAGE:33575"
/clone_lib="Soares infant brain INIB"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lfamid BR; Site_1: Not I; Site_2: Hind III; lstr strand cDNA was primed with a Not I - oligo(dT) primer [5'
AACTGGAAGAAATTCGCGCGCAGCAATTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors
(Pharmacia), digested with Not I and directionally cloned
into the Not I and Hind III sites of the Lfamid BR vector.
Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."
human.
129 a 58 c 4 others
BASE COUNT 129 a 58 c 150 t
ORIGIN

Query Match 87.0%; Score 17.4; DB 21; Length 407;
Best Local Similarity 94.7%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTCCAT 19
|||||
Db 338 TTACATATGAGCCTCCAT 320

RESULT 2
D55860/c 360 bp mRNA EST 31-AUG-1995
LOCUS HUM404C08B Clontech human fetal brain polyA+ mRNA (#5535) Homo
DEFINITION sapiens cDNA clone GEN-404C08 5', mRNA sequence.
ACCESSION D55860
NID g970260
VERSION D55860.1 GI:970260
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 360)

AUTHORS Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y., Maekawa, H., Shin, S. and Nakamura, Y.
Fujiwara et al. (1995)
Unpublished (1995)
On Sep 21, 1992 this sequence version replaced gi:279304.

TITLE
JOURNAL
COMMENT Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035
Insert Length: 802 Std Error: 0.00
High quality sequence stop: 353.
Location/Qualifiers
1..360
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-404C08"
/clone_lib="Clontech human fetal brain polyA+ mRNA
(#5535)"
96 a 49 c 63 g 152 t
BASE COUNT 96 a 49 c 63 g 152 t
ORIGIN

Query Match 84.0%; Score 16.8; DB 23; Length 360;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTCCATG 20
|||||
Db 313 TTACATATGAGCCTTCATG 294

RESULT 3
N29325
LOCUS Yv84cl1.s1 Soares placenta_8to9weeks_2NBHP8to9w Homo sapiens cDNA
DEFINITION Clone IMAGE:258932 3', mRNA sequence.
ACCESSION N29325
NID g1147845
VERSION N29325.1 GI:1147845
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 476)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)
On May 18, 1995 this sequence version replaced gi:810971.

TITLE
JOURNAL
COMMENT Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 278
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: ml3 -40 forward
High quality sequence stop: 278.
Location/Qualifiers
1..476

```

/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes=Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCTCATTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT      183 a      94 c      69 g      142 t
ORIGIN

Query Match      84.0%; Score 16.8; DB 39; Length 488;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 TTACATATGAGCCTTCATG 20
        |||||
Db       80 TTACATATGAGCTTCATG 99
        |||||

RESULT      5
AA908884    493 bp      mRNA      EST      23-JUN-1998
LOCUS      OM51402.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1550715 3'
DEFINITION similar to SW:YN8H_YEAST P53729 HYPOTHETICAL 48.1 KD PROTEIN IN
SEC12-S5K2 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION  AA908884
NID         G3048289
VERSION     AA908884.1 GI:3048289
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 493)
AUTHORS    NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished (1997)
COMMENT     On Jan 17, 1998 this sequence version replaced gi:2044779.

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 905 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 467.
Location/Qualifiers
1. 493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1550715"
/clone_lib="NCI-CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH108"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified

polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7n3 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 193 a 94 c 67 g 139 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 40; Length 493;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
||||| ||||| ||||| |||||

Db 78 TTACATATGAGACTTTCATG 97
||||| ||||| ||||| |||||

RESULT 6
LOCUS AI130847 598 bp mRNA EST 27-OCT-1998
DEFINITION qc14a10.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone IMAGE:1709562 3' similar to SW:YN8H_YEAST P53729 HYPOTHETICAL 48.1 KD PROTEIN IN SEC12-SSK2 INTERGENIC REGION. ; mRNA sequence.

ACCESSION AI130847
NID q3600863
VERSION AI130847.1 GI:3600863
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 598)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 9, 1998 this sequence version replaced gi:930548.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 771 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 440.

FEATURES
source
1..598
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1709562"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTITTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NBHL19W."

BASE COUNT 211 a 123 c 88 g 175 t 1 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 42; Length 598;
Best Local Similarity 90.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
||||| ||||| ||||| |||||

Db 30 TTACATATGAGACTTTCATG 99
||||| ||||| ||||| |||||

RESULT 7
LOCUS AI248140 567 bp mRNA EST 01-DEC-1998
DEFINITION qh73b12.x1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA clone IMAGE:1850495 3' similar to SW:YN8H_YEAST P53729 HYPOTHETICAL 48.1 KD PROTEIN IN SEC12-SSK2 INTERGENIC REGION. ; mRNA sequence.

ACCESSION AI248140
NID q3843537
VERSION AI248140.1 GI:3843537
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 567)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 19, 1998 this sequence version replaced gi:2151692.

CONTACT: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 772 Std Error: 0.00
Seq primer: -400P from Gibco
High quality sequence stop: 354.

FEATURES
source
1..567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="16p13.3"
/clone="IMAGE:1850495"
/clone_lib="Soares_fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGGAAGAATTAATAAGATCTTTTITTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 205 a 109 c 82 g 169 t 2 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 44; Length 567;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
||||| ||||| ||||| |||||

Db 83 TTACATATGAGACTTTCATG 102
||||| ||||| ||||| |||||

RESULT 8
D52258/c
LOCUS
DEFINITION HUM072G05B Clontech human fetal brain polyA+ mRNA EST 14-SEP-1995
sapiens CDNA clone GEN-072G05 5', mRNA sequence.
ACCESSION D52258
NID 9952494
VERSION D52258.1 GI:952494
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 379)
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
Maeawa,H., Shin,S. and Nakamura,Y.
Fujiwara et al. (1995)
JOURNAL Unpublished (1995)
COMMENT On May 8, 1995 this sequence version replaced gi:801457.

Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co.,Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035
Insert Length: 572 Std Error: 0.00
High quality sequence stop: 328.

FEATURES
source
1. .379
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-072G05"
/clone_lib="Clontech human fetal brain polyA+ mRNA
(#6535)"
BASE COUNT 100 a 51 c 62 g 159 t 7 others
ORIGIN
|||||
Query Match 82.0%; Score 16.4; DB 23; Length 379;
Best Local Similarity 85.0%; Pred. No. 58;
Matches 1; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 1 TTACATATGAGCCTTCATG 20
|||||
Db 314 TTACATATGAGCCTTCATG 295

RESULT 9
D53341/c
LOCUS
DEFINITION HUM106D02B Clontech human fetal brain polyA+ mRNA EST 19-SEP-1995
sapiens CDNA clone GEN-106D02 5', mRNA sequence.
ACCESSION D53341
NID 9955238
VERSION D53341.1 GI:955238
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 402)
AUTHORS Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
Maeawa,H., Shin,S. and Nakamura,Y.
Fujiwara et al. (1995)
JOURNAL Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:716975.
Contact: Tsutomu Fujiwara

Otsuka GEN Research Institute
Otsuka Pharmaceutical Co.,Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.

FEATURES
source
1. .402
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GEN-106D02"
/clone_lib="Clontech human fetal brain polyA+ mRNA
(#6535)"
BASE COUNT 115 a 53 c 67 g 164 t 3 others
ORIGIN
|||||
Query Match 82.0%; Score 16.4; DB 23; Length 402;
Best Local Similarity 85.0%; Pred. No. 59;
Matches 17; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Oy 1 TTACATATGAGCCTTCATG 20
|||||
Db 313 TTACATATGAGACTKTCATG 294

RESULT 10
H90734/c
LOCUS
DEFINITION H90734 447 bp mRNA EST 28-NOV-1995
yu84a04.s1 Soares fetal liver spleen lNFLS Homo sapiens cDNA clone
IMAGE:240462 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION H90734
NID G1081164
VERSION H90734.1 GI:1081164
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 447)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
COMMENT On Apr 14, 1993 this sequence version replaced gi:838025.

Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 809
High quality sequence stops: 369
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 809 Std Error: 0.00
Seq primer: Promega -2lm13
High quality sequence stop: 369.

FEATURES
source
1. .447
/organism="Homo sapiens"
/db_xref="GDB:3789595"
/db_xref="taxon:9606"
/clone="IMAGE:240462"
/clone_lib="Soares fetal liver spleen lNFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"

/note="Organ: Liver and Spleen; Vector: pT773D (Pharmacia) with a modified polylinker; Site_1: Pac I - oligo(dR) primer 1st strand cDNA was primed with a Pac I - oligo(dR) primer [5', ACTGTGAGAAATTAATAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library went through two more rounds and M.Fatima Ronaldo "

```

BASE COUNT      129 a      96 c      104 g      101 t      17 others
ORIGIN

Query Match      79.0%      score 15.8;  DB 24;  Length 447;
Best Local Similarity 89.5%      Pred. No. 1.2e+02;
Matches 17;  Conservative 2;  Mismatches 2;  Indels 0;  Gaps 0;

```

Qy 2 TACATATGAGCCTTCCATG 20
||||| ||||| ||||| |||||
Db 280 TACATTGAGGCTTCCATG 262

[illegible]

ACCESSION	H95824	
NID	91108966	
VERSION	H95824.1	GI:1108966
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 424)
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Hollman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevisakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE	The WashU-Merck EST Project
JOURNAL	Unpublished (1995)
COMMENT	On Jan 25, 1995 this sequence version replaced qi:637882.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: estevan@wustl.edu
High quality sequence stops: 315
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 804 Std Error: 0.00
Seq primer: ml3 -40 forward.

```

FEATURES
source
Location/Qualifiers
1. .424
/organism="Homo sapiens"
/db_xref="gdb:3792393"
/db_xref="taxon:9606"
/clone="IMAGE:243260"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(AT) primer [5' - AACTGGAGATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac

```

and Eco RI sites of the modified pT73 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

125 a	87 c	94 g	117 t	1 others

```
Query Match          79.0%; Score 15.8; DB 24; Length 424;
Best Local Similarity 89.5%; Pred. NO. 1.2e+02;
Matches 17: Conservative 0; Mismatches 2; Indels 0; Gaps
```

Qy 2 TACATATGAGCCTTCCATG 20
||||| ||||| ||||| |||||
Db 296 TACATTTGAGGCTTCCATG 278

RESULT	12
AA197351	
LOCUS	
DEFINITION	630 bp mRNA EST 19-FEB-1997
	muscle08.r1 Soares 2NbMt Mus musculus cDNA clone IMAGE:640066 5'
	mRNA sequence.

ACCESSION AA197351
NID q1792993
VERSION AA197351.1 CI:1792993
KEYWORDS EST.

SOURCE	ORGANISM
house mouse.	<i>Mus musculus</i>
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
1 (bases 1 to 630)
Marra,N., Kuller,L., Allen,M., Bowles,M., Dietrich,N., Dubuque
Geisel,S., Kuback,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE	The WashU-HHMI Mouse EST Project
JOURNAL	Unpublished (1996)
COMMENT	On Sep 12, 1996 this sequence version replaced qi:1393441.

Contact: Marra M/Mouse EST Project
WASHU-HHMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:397058

seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 507.

```
FEATURES
source
    1..630
    Location/Qualifiers
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone_image="IMAGE:640066"
        /clone_lib="Soares 2NbMT"
        /sex="male"
        /tissue_type="Thymus"
        /dev_stage="4 weeks"
        /lab_book="nu10B"
```

BASE COUNT
156 a 130 c 157 q 187 t
/lap_035c= drabv
[note=vector: p773D-Pac (pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand
was primed with Not I - oligo(dT) primer [5'.
TGTTCACCTGAAGTGGAGCCGCCGCTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adapted
(Pharmacia), digested with Not I and cloned into the N
and Eco RI sites of the modified p773 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
rounds of normalization, and was constructed by Bento
Soares and M.Patrina Bonaldo."

○

[illegible]

Mon Sep 20 14:16:38 1999

```

RESULT 15
AA377278/c 372 bp mRNA EST 21-APR-1997
LOCUS EST89818 Small intestine II Homo sapiens cDNA 3' end, mRNA
DEFINITION sequence.
ACCESSION AA377278
NID 92029596
VERSION AA377278.1 GI:2029596
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE *Rapid cDNA sequencing (expressed sequence tags) from a
directionally cloned human infant brain cDNA library
JOURNAL Nature Genet. 4, 373-380 (1993)
MEDLINE 94004965
COMMENT On Sep 12, 1996 this sequence version replaced gi:1288980.
Other ESTs: EST89819 TH2137556
Contact: Kerlavage, AF
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (http://www.tigr.org/tcdb/hgi/hgi.html)
Seq primer: M13-21.
FEATURES
    source
        1..372
            /organism="Homo sapiens"
            /db_xref="ATCC (ihost):181679"
            /db_xref="taxon:9606"
            /clone_lib="Small intestine II"
            /dev_stage="adult"
            /note="Organ: small intestine; Vector: pBluescript SK-;
                Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 115 a 68 c 72 g 114 t 3 others
ORIGIN

```

```

Query Match 79.0%; Score 15.8; DB 32; Length 372;
Best Local Similarity 89.5%; Pred. NO. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 TACATATGAGCCTTCATG 20
    ||||| ||||| ||||| |||||
Db 314 TACATTGAGGCTTCATG 296

```

Search completed: September 18, 1999, 06:47:49
Job time: 3666 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 15:49:29 ; Search time 436.05 Seconds
(without alignments)
196.923 Million cell updates/sec

Title: US-09-037-472-9

Perfect score: 27

Sequence: 1 AAGCTTGTCTTACCACCTGAAGTAGGC 27

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*

1: gb_ba1:*
2: gb_ba2:*
3: gb_om:*
4: gb_ov:*
5: gb_pat:*
6: gb_ph:*
7: gb_pl1:*
8: gb_pl2:*
9: gb_pr1:*
10: gb_pr2:*
11: gb_pr3:*
12: gb_ro:*
13: gb_st:*
14: gb_sts:*
15: gb_sy:*
16: gb_un:*
17: gb_vi:*
18: em_fun:*
19: em_htg:*
20: em_hum1:*
21: em_hum2:*
22: em_in:*
23: em_om:*
24: em_of:*
25: em_ov:*
26: em_pat:*
27: em_ph:*
28: em_pl:*
29: em_ro:*
30: em_sts:*
31: em_sy:*
32: em_un:*
33: em_vi:*
34: gb_htg1:*
35: gb_htg2:*
36: gb_in1:*
37: gb_in2:*
38: em_ba1:*
39: em_ba2:*
40: em_hum3:*
41: em_hum4:*
42: gb_pr4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	22	81.5	22	5	I73225 Sequence 1

2	22	81.5	11970	10	HS11IAG	X03833 Human gene
3	18.6	68.9	145614	34	AC006804	AC006804 Caenorhab
4	18.6	68.9	145614	34	AC006872	AC006872 Caenorhab
5	18.6	68.9	244239	34	AC006880	AC006880 Caenorhab
6	18.6	68.9	151528	35	AC006447	AC006447 Mus muscu
7	18.4	68.1	170952	9	HS01N12	AL022170 Homo sapi
8	18.2	67.4	103574	10	HSAC002115	AC002115 Human DNA
9	18	66.7	1910	4	CHKLNKPA	M35036 Chicken car
10	18	66.7	14313	7	NEUROIDHC	L31504 Neurospora
11	18	66.7	139166	11	AC003029	AC003029 Homo sapi
12	18	66.7	151795	34	HS856G1	AL033381 Homo sapi
13	18	66.7	51050	36	AC004340	AC004340 Drosophi
14	18	66.7	85139	37	AC001657	AC001657 Drosophi
15	18	66.7	196672	37	AC004767	AC004767 Drosophi
16	17.6	65.2	16094	1	MTV038	AL021933 Mycobacte
17	17.6	65.2	783	3	AF030474	AF030474 Gazella s
18	17.6	65.2	783	3	AF030475	AF030475 Gazella c
19	17.6	65.2	783	3	AF030476	AF030476 Gazella l
20	17.6	65.2	783	3	AF030477	AF030477 Gazella s
21	17.6	65.2	783	3	AF030478	AF030478 Gazella s
22	17.6	65.2	783	3	AF030479	AF030479 Gazella b
23	17.6	65.2	783	3	AF030480	AF030480 Gazella b
24	17.6	65.2	783	3	AF030481	AF030481 Gazella s
25	17.6	65.2	783	3	AF030483	AF030483 Gazella d
26	17.6	65.2	783	3	AF030484	AF030484 Gazella d
27	17.6	65.2	783	3	AF030485	AF030485 Gazella g
28	17.6	65.2	783	3	AF030487	AF030487 Gazella g
29	17.6	65.2	783	3	AF030488	AF030488 Gazella g
30	17.6	65.2	783	3	AF030489	AF030489 Gazella g
31	17.6	65.2	783	3	AF030489	AC006812 Caenorhab
32	17.6	65.2	16034	34	AC006812	AC006812 Caenorhab
33	17.6	65.2	16034	34	AC006812	AC001232 Homo sapi
34	17.6	65.2	90143	34	HSAC001232	AC007255 Homo sapi
35	17.6	65.2	171336	35	AC007255	AB002762 Entamoeba
36	17.6	65.2	473	36	AB002762	X14386 Astasia lon
37	17.4	64.4	6081	7	ALCRDNA	297209 S.pombe chr
38	17.4	64.4	37000	7	SPAC19G12	AF038122 Podospora
39	17.4	64.4	3057	8	AF038122	D49396 Human mRNA
40	17.4	64.4	1542	9	HUMAOPI	AC004823 Homo sapi
41	17.4	64.4	115973	11	AC004823	AC006398 Homo sapi
42	17.4	64.4	92611	11	AC006398	AC006466 Homo sapi
43	17.4	64.4	184514	34	LMFL1156	AL034390 Leishmani
44	17.4	64.4	39897	34	LMFL1156	AC007778 Homo sapi
45	17.4	64.4	167713	35	AC007778	

ALIGNMENTS

RESULT	1	LOCUS	I73225	Sequence 1 from patent US 5686246.	22 bp	DNA	PAT	23-DEC-1997
I73225		DEFINITION	I73225					
		ACCESSION	I73225					
		NID	g3009364					
		VERSION	I73225.1	GI:3009364				
		KEYWORDS						
		SOURCE	Unknown.					
		ORGANISM	Unknown.					
		REFERENCE	Unclassified.					
		AUTHORS	Kornman, K. S. and Duff G.W.					
		TITLE	Detecting genetic predisposition to periodontal disease					
		JOURNAL	Patent: US 5686246-A 1 11-NOV-1997;					
		FEATURES	Location/Qualifiers					
		source	1..22					
		BASE COUNT	5 a	7 c	4 g	6 t		
		ORIGIN						

Query Match 81.5%; Score 22; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2;

```

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 6 TGTCTACCACTGAAGTAGGC 27
Db 1 TGTCTACCACTGAAGTAGGC 22

RESULT 2
HSILIAG          11970 bp      DNA      PRI      24-APR-1993
LOCUS            Human gene for interleukin 1 alpha (IL-1 alpha).
DEFINITION       X03833
ACCESSION        33785
NID              33785
VERSION          X03833.1 GI:33785
KEYWORDS         Alu repetitive sequence; interleukin 1 alpha; inverted repeat;
SOURCE           human.
ORGANISM         Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE        1 (bases 1 to 11970)
AUTHORS          Furutani,Y., Notake,M., Fukui,T., Ohue,M., Nomura,H., Yamada,M. and
                  Nakamura,S.
TITLE            Complete nucleotide sequence of the gene for human interleukin 1
                  alpha
JOURNAL          Nucleic Acids Res. 14 (8), 3167-3179 (1986)
MEDLINE          86205226
ERRATUM:[[published erratum appears in Nucleic Acids Res 1986 Jun
REMARK          25:14(12):5124]]
COMMENT          Data kindly reviewed (10-NOV-1986) by Y. Furutani.
FEATURES         Location/Qualifiers
                     source          1..11970
                                     /organism="Homo sapiens"
                                     /db_xref="taxon:9606"
                     misc_feature    254..393
                                     /note="Alu repetitive sequence"
                     repeat_region    1375..1382
                                     /note="direct repeat 1"
                     misc_signal      1375..1390
                                     /note="pot. transcription activator (seq. homolog. to
                                     adenovirus 2 major late promoter transcription factor
                                     (MLTF) binding site)"
                     repeat_unit       1375..1382
                                     /note="inverted repeat A"
                     repeat_unit       1383..1390
                                     /note="inverted repeat A'"
                     repeat_region     1383..1390
                                     /note="direct repeat 1"
                     TATA_signal       1407..1413
                     prim_transcript  1438..11643
                     exon              1438..1488
                                     /number=1
                     mRNA             join(1438..1488,2153..2207,3166..3214,4103..4325,
                                     5262..6432,7815..7939,10290..11643)
                     intron           2153..2207
                                     /number=1
                     exon              2153..2207
                                     /number=2
                     CDS               join(2161..2207,3166..3214,4103..4325,6262..6432,
                                     7815..7939,10290..10490)
                                     /codon_start=1
                                     /product="IL-1-alpha"
                                     /protein_id="CAA27448.1"
                                     /db_xref="PID:g33786"
                                     /db_xref="GI:33786"
                                     /db_xref="SWISS-PROT:P01583"
                                     /translation="MAKVPDMFEDLNKCYSENEEDSSSDHLSLNQKSPHYHYSGPLH
                                     ECDQSVLSISSETSKLTFKESMVVATNGKVKRRLSLSQSITDDLEAIAN
                                     DSEELIKPRSSPFSLSNVKYNFMRIKYEFILNALNQSIIRANDQVLTAAHLN
                                     DEAVKFDMGAYKSKDKDKITVILRIKSTQLVTAODEOPVLLKEMPEIPKTIYGE
                                     TLLFFWEHFGTKNYFTVAHNPFLFIATKQDIWVCLAGGPPSITDFQILENOA"
                                     2208..3165
                     intron           /number=2

```

```

exon              3166..3214
                  /number=3
intron            3215..4102
                  /number=3
exon              4103..4325
                  /number=4
intron            4326..6261
                  /number=4
misc_feature      4893..5174
                  /note="Alu repetitive sequence"
exon              5262..6432
                  /number=5
intron            6433..7814
                  /number=5
misc_feature      7695..7744
                  /note="poly [dA-dC] tract"
exon              7815..7939
                  /number=6
intron            7940..10289
                  /number=6
misc_feature      8466..8483
                  /note="poly[dA-dC] tract"
repeat_region     8912..9137
                  /note="5 x 46 bp repeat"
misc_feature      9770..9806
                  /note="poly [dT-dG] tract"
exon              10290..11643
                  /number=7
polyA_site        11643
                  /note="polyA site"
misc_feature      11863..11970
                  /note="Alu repetitive sequence"
BASE COUNT       3708 a 2489 c 2226 g 3547 t
ORIGIN

Query Match      81.5%; Score 22; DB 10; Length 11970;
Best Local Similarity 100.0%; Pred.No.2.1;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TGTCTACCACTGAAGTAGGC 27
    |||||
Db 471 TGTCTACCACTGAAGTAGGC 492

RESULT 3
AC006804
LOCUS            AC006804.2 GI:4309782
DEFINITION       HTG; HTGS.PHASE1.
KEYWORDS         Caenorhabditis elegans.
SOURCE           Caenorhabditis elegans.
ORGANISM         Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
                  Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.
                  Waterston,R.H.
                  The sequence of Caenorhabditis elegans clone
REFERENCE        1 (bases 1 to 145614)
AUTHORS          Waterston,R.H.
JOURNAL          Unpublished
TITLE            Direct Submission
REFERENCE        2 (bases 1 to 145614)
AUTHORS          Waterston,R.H.
JOURNAL          Submitted (23-FEB-1999) Genome Sequencing Center, Washington
                  University School of Medicine, 444 Forest Park Parkway, St. Louis,
                  MO 63108, USA
COMMENT          On Mar 1, 1999 this sequence version replaced gi:4263134.
                  * NOTE: This is a 'working draft' sequence. It currently
                  * consists of 5 contigs. The true order of the pieces
                  * is not known and their order in this sequence record is
                  * arbitrary. Gaps between the contigs are represented as

```

* runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 19062: contig of 19062 bp in length
 * 19063 19071: gap of unknown length
 * 19072 27630: contig of 8559 bp in length
 * 27631 27639: gap of unknown length
 * 27640 54395: contig of 26756 bp in length
 * 54396 54404: gap of unknown length
 * 54405 56414: contig of 2010 bp in length
 * 56415 56423: gap of unknown length
 * 56424 145614: contig of 89191 bp in length.

FEATURES

source

1. 145614
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="Y53G88"

BASE COUNT 46966 a 25460 c 25159 g 47086 t 943 others
 ORIGIN

Query Match 68.9%; Score 18.6; DB 34; Length 145614;
 Best Local Similarity 84.0%; Pred. No. 79;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGCTTGTTCTACCACTGAACTAG 25

Db 53675 ATGCTTCTTCTACCACTGAAATG 53699

RESULT 4

AC006872

LOCUS

DEFINITION AC006872 145614 bp DNA HTG 26-FEB-1999
 Caenorhabditis elegans clone Y53G8Y, WORKING DRAFT SEQUENCE, 5
 unordered pieces.

ACCESSION AC006872

NID 94309781

VERSION AC006872.2 GI:4309781

KEYWORDS HTG; HTGS-PHASE1.

SOURCE Caenorhabditis elegans.

ORGANISM

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 Waterston.R.H.

REFERENCE 1 (bases 1 to 145614)

TITLE The sequence of Caenorhabditis elegans clone

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 145614)

Waterston.R.H.

Direct Submission

Submitted (24-FEB-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT On Mar 1, 1999 this sequence version replaced GI:4263476.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 19062: contig of 19062 bp in length

* 19063 19071: gap of unknown length

* 19072 27630: contig of 8559 bp in length

* 27631 27639: gap of unknown length

* 27640 54395: contig of 26756 bp in length

* 54396 54404: gap of unknown length

* 54405 56414: contig of 2010 bp in length

* 56415 56423: gap of unknown length

* 56424 145614: contig of 89191 bp in length.

Location/Qualifiers

FEATURES

source

1. 145614
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="Y53G8Y"

BASE COUNT 46966 a 25460 c 25159 g 47086 t 943 others
 ORIGIN

Query Match 68.9%; Score 18.6; DB 34; Length 145614;
 Best Local Similarity 84.0%; Pred. No. 79;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGCTTGTTCTACCACTGAACTAG 25

Db 53675 ATGCTTCTTCTACCACTGAAATG 53699

RESULT 5

AC006880

LOCUS

DEFINITION AC006880 244239 bp DNA HTG 24-FEB-1999
 Caenorhabditis elegans clone Y55D5, WORKING DRAFT SEQUENCE, 4
 unordered pieces.

ACCESSION AC006880

NID 94263468

VERSION AC006880.1 GI:4263468

KEYWORDS HTG; HTGS-PHASE1.

SOURCE Caenorhabditis elegans.

ORGANISM

Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
 Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
 Waterston.R.H.

REFERENCE 1 (bases 1 to 244239)

TITLE The sequence of Caenorhabditis elegans clone

JOURNAL Unpublished

AUTHORS 2 (bases 1 to 244239)

Waterston.R.H.

Direct Submission

Submitted (24-FEB-1999) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

COMMENT

* NOTE: This is a 'working draft' sequence. It currently

* consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 9998: contig of 9998 bp in length

* 9999 10012: gap of unknown length

* 10013 19150: contig of 9138 bp in length

* 19151 19164: gap of unknown length

* 19165 88195: contig of 69031 bp in length

* 88196 88209: gap of unknown length

* 88210 244239: contig of 156030 bp in length.

Location/Qualifiers

source

1. 244239
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /clone="Y55D5"

BASE COUNT 78698 a 43791 c 42693 g 79015 t 42 others
 ORIGIN

Query Match 68.9%; Score 18.6; DB 34; Length 244239;
 Best Local Similarity 84.0%; Pred. No. 79;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AGCTTGTTCTACCACTGAACTAG 25

Db 87475 ATGCTTCTTCTACCACTGAAATG 87499

```

RESULT 6
AC006447
LOCUS      AC006447      151528 bp      DNA      HTG      09-JUN-1999
DEFINITION Mus musculus, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION  AC006447
NID         95030453
VERSION     AC006447.6 GI:5030453
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 151528)
AUTHORS     Wu.H., Yao.Z., McDermid,H. and Roe,B.A.
TITLE       Mus musculus Chromosome 6 BAC Clone 67d14
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 151528)
AUTHORS     Wu.H., Yao.Z., McDermid,H. and Roe,B.A.
TITLE       Direct Submission
JOURNAL     Submitted (28-JAN-1999) Department of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT     On Jun 9, 1999 this sequence version replaced gi:4580404.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 5480: contig of 5480 bp in length
            * 5481 5555: gap of unknown length
            * 5556 12748: contig of 7193 bp in length
            * 12749 12823: gap of unknown length
            * 12824 37024: contig of 24201 bp in length
            * 37025 37029: gap of unknown length
            * 37100 82030: contig of 44930 bp in length
            * 82030 82104: gap of unknown length
            * 82105 151528: contig of 69424 bp in length.
FEATURES             Location/Qualifiers
     source           1..151528
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
BASE COUNT          43172 a 33398 c 33443 g 41215 t      300 others
ORIGIN
Query Match      58.9%; Score 18.6; DB 35; Length 151528;
Best Local Similarity 84.0%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Oy 1 AAGCTTGTTCTACCACCTGAACCTAG 25
      |||||  |||||  |||||  |||||
Db 98144 AAGCTTGTTGTCACCTACCTAG 98168

RESULT 7
HS501N12/c
LOCUS      HS501N12      170952 bp      DNA      PRI      24-JUL-1998
DEFINITION Homo sapiens DNA sequence from clone 501N12 on chromosome
            6p22.1-22.3. Contains a gene almost identical to four genes of
            unknown function, a pseudogene, three (pseudo?) genes similar to
            genes of unknown function, an unknown gene similar to a rat EST, a
            PX19 Like pseudogene and another unknown gene. Contains ESTs, STSs
            and GSSs, complete sequence.
ACCESSION    AL022170
NID          93281976
VERSION      AL022170.1 GI:3281976
KEYWORDS     HTG; PX19.
SOURCE      human.
ORGANISM     Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

```

```

REFERENCE
AUTHORS     Tubby,B.
TITLE       Direct Submission
JOURNAL     Submitted (06-JUL-1998) E-mail enquires: humquery@sanger.ac.uk
COMMENT     Clone requests: clonerequest@sanger.ac.uk
            On Jul 1, 1998 this sequence version replaced gi:2980811.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above. This sequence is the
            entire insert of clone 501N12. This sequence has been finished
            according to sequence map criteria as follows. An attempt is made
            to resolve all sequencing problems, such as compressions and
            repeats, but not necessarily within known annotated human repeat
            sequence elements (e.g. Alu). Where the sequence is ambiguous,
            there is an annotation using the 'unsure' feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed in collaboration by the Sanger
            Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6/ 501N12 is from the library RPC13
            constructed at the Roswell Park Cancer Institute by the group of
            Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/ VECTOR: pcYPAC2.
FEATURES             Location/Qualifiers
     source           1..170952
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="501N12"
                     /map="p22.1-22.3"
                     /clone_lib="RPC13"
                     444..516
                     /note="MER5A repeat: matches 101..188 of consensus"
                     complement(832..1100)
                     /note="AluJb repeat: matches 280..1 of consensus"
                     complement(1341..1477)
                     /note="MIR repeat: matches 237..85 of consensus"
                     2942..3479
                     /note="L1 repeat: matches 1432..1968 of consensus"
                     4030..4359
                     /note="AluY repeat: matches 1..296 of consensus"
                     complement(4554..4853)
                     /note="AluJo repeat: matches 301..1 of consensus"
                     6829..34852
                     /gene="dJ501N12.1"
                     join(<6829..6895,15857..15985,17949..18080,28912..29087,
                     32011..32098,34703..34852)
                     /gene="dJ501N12.1"
                     /note="match: cDNAs Y10518 Y10519 283950 Y10508"
                     /codon_start=2
                     /evidence=not_experimental
                     /product="dJ501N12.1"
                     /protein_id="CAA18156.1"
                     /db_xref="PID:e1312796"
                     /db_xref="PID:g3355565"
                     /db_xref="GI:3355565"
                     /db_xref="SPTREMBL:O75660"
                     /translation="PKYLQEQAYRLQOGLIVSTTQMDIRICVKVODHLNLSRNCGGD
                     ALQEDLKAERLMROAKSKTLLPNLYHVGGSWAGAGSLSPQETLSMAGEVTR
                     VDEQLKALLESMDVAENLCPNVMKKAHRODLIHASTEKIIPRTFVKNVLLQSG
                     IDILNKISEVKLTVASDFNCLPVMKKAHRODLIHASTEKIIPRTFVKNVLLQSG
                     AEEKPVKRSIITVEELTEIERLEDLDTCM"
     prim_transcript  7271..7658
                     /gene="dJ501N12.1"
                     /note="match: EST AA526622"
                     7659..7776
                     /note="AluYb8 repeat: matches 191..308 of consensus"
                     8239..8475
                     /note="AluSg repeat: matches 1..239 of consensus"

```

```
misc_feature      complement(8536..8952)
                  /gene="dJ501N12.1"
                  /note="match: GSS B37778"
repeat_region     complement(9800..10072)
                  /note="AluX repeat: matches 273..1 of consensus"
repeat_region     complement(10099..10222)
                  /note="AluSp repeat: matches 303..179 of consensus"
repeat_region     complement(11503
11294..11503)
                  /note="MIR repeat: matches 4..214 of consensus"
prim_transcript   complement(12237..12525)
                  /gene="dJ501N12.1"
                  /note="match: EST AA639318"
repeat_region     complement(12599..12717)
                  /note="MIR2 repeat: matches 146..28 of consensus"
repeat_region     complement(15448..15519)
                  /note="MIR2 repeat: matches 74..145 of consensus"
misc_feature      complement(15866..16333)
                  /gene="dJ501N12.1"
                  /note="match: GSS AQ016761"
repeat_region     complement(16512..16606)
                  /note="MER21B repeat: matches 790..703 of consensus"
repeat_region     complement(16545..16637)
                  /note="MER39 repeat: matches 676..582 of consensus"
repeat_region     complement(16603..17427)
                  /note="MER21B repeat: matches 792..4 of consensus"
repeat_region     complement(17480..17647)
                  /note="AluSp repeat: matches 136..303 of consensus"
repeat_region     complement(19382..19433)
                  /note="L1HS repeat: matches 578..526 of consensus"
repeat_region     complement(19556..19960)
                  /note="AluX repeat: matches 300..1 of consensus"
repeat_region     complement(22648
22609..22648)
                  /note="20 copies 2 mer ac 100% conserved"
repeat_region     complement(22914..23103)
                  /note="MER3 repeat: matches 209..1 of consensus"
repeat_region     complement(24346
24161..24346)
                  /note="L1 repeat: matches 4042..4231 of consensus"
repeat_region     complement(24589
24420..24589)
                  /note="AluJo repeat: matches 1..302 of consensus"
repeat_region     complement(25083
24771..25083)
                  /note="AluSq repeat: matches 2..303 of consensus"
repeat_region     complement(26904
26805..26904)
                  /note="MIR repeat: matches 35..134 of consensus"
repeat_region     complement(27877..28054)
                  /note="MIR repeat: matches 260..77 of consensus"
repeat_region     complement(28238
28097..28238)
                  /note="MIR repeat: matches 71..201 of consensus"
misc_feature      complement(28312..28688)
                  /gene="dJ501N12.1"
                  /note="match: GSS AQ020591"
repeat_region     complement(30856..30913)
                  /note="MER5A repeat: matches 111..57 of consensus"
repeat_region     complement(33402..33698)
                  /note="AluX repeat: matches 301..1 of consensus"
repeat_region     complement(38477..38699)
                  /note="MIR repeat: matches 253..2 of consensus"
repeat_region     complement(40249..40548)
                  /note="AluSq repeat: matches 299..1 of consensus"
repeat_region     complement(41636..41692)
                  /note="MIR2 repeat: matches 146..87 of consensus"
repeat_region     complement(42990..43016)
                  /note="MIR2 repeat: matches 98..124 of consensus"
repeat_region     complement(43860
43567..43860)
                  /note="AluSc repeat: matches 1..298 of consensus"
repeat_region     complement(44723
44628..44723)
                  /note="MIR2 repeat: matches 30..134 of consensus"
repeat_region     complement(45009..45166)
                  /note="MIR repeat: matches 180..21 of consensus"
repeat_region     complement(47974..48264)
                  /note="AluY repeat: matches 301..2 of consensus"
mRNA              complement(48410..49073)
                  /gene="dJ501N12.2"
                  /note="match: cDNA Y10205"

/pseudo
/evidence-not_experimental
complement(48410..49073)
/gene="dJ501N12.2"
49166..50348
/note="MER42c repeat: matches 348..1536 of consensus"
50415..50724
/note="AluJo repeat: matches 2..302 of consensus"
complement(50998..51085)
/note="MIR2 repeat: matches 145..58 of consensus"
complement(52855..52797)
/note="L1ME2 repeat: matches 699..554 of consensus"
complement(52991..53050)
/note="MIR repeat: matches 234..175 of consensus"
complement(60954..60997)
/note="MER34 repeat: matches 524..480 of consensus"
complement(61036..61200)
/note="MER39 repeat: matches 409..251 of consensus"
61430..61537
/note="MIR2 repeat: matches 32..142 of consensus"
complement(62029..62337)
/note="AluJo repeat: matches 301..6 of consensus"
62452..62815
/note="MER39 repeat: matches 48..409 of consensus"
62854..62898
/note="MER34 repeat: matches 480..525 of consensus"
63567..65020
/note="SVA repeat: matches 4..1372 of consensus"
complement(65545..65881)
/note="match: STS G05436"
complement(65917..66232)
/note="MER2 repeat: matches 345..9 of consensus"
complement(67491..67789)
/note="AluJo repeat: matches 300..1 of consensus"
complement(68357..68545)
/note="L1 repeat: matches 3768..3579 of consensus"
complement(70789..71348)
/gene="dJ501N12.3"
/note="possibly a pseudogene; match: cDNAs Y10202 283943"
/evidence-not_experimental

Query Match      68.1%; Score 18.4; DB 9; Length 170952;
Best Local Similarity 95.0%; Pred No. 98;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 6 TGTTCTACCACTGAAGTAG 25
    ||| ||||| ||||| |||||
Db 38328 TGCTCTACCACTGAAGTAG 38309

RESULT 8
HSAC002115 LOCUS
DEFINITION HSAC002115 103574 bp DNA PRI 30-JUN-1997
            Human DNA from overlapping chromosome 19 cosmids R31396, F25451,
            and R31076 containing COX8 and UPK8, genomic sequence, complete
            sequence.
ACCESSION AC002115
NID 92098573
VERSION AC002115.1 GI:2098573
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 103574)
AUTHORS Lamerdin,J.E., McCready,P.M., Adamson,A.W., Burkhardt-Schultz,K.,
            Garcia,E., Kyle,A., Ramirez,M., Stillwagen,S., Ganes,J.,
            Danganan,L., Bruce,R., Quan,G., Montgomery,M., Ow,D., Kobayashi,A.,
            Olsen,A.O. and Carrano,A.V.
            Sequence analysis of a 1 Mb region in human 19q13.1
            Unpublished (1997)
REFERENCE 2 (bases 1 to 103574)
AUTHORS Lamerdin,J.E.
```

```

TITLE      Direct Submission
JOURNAL    Submitted (13-MAY-1997) Human Genome Center, Lawrence Livermore
           National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
COMMENT    R31396 from 1- 36,162; F25451 from 25,661-55,793; R31076 from
           66,237- 103,574
FEATURES   Location/Qualifiers
           source          1..103574
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="R31396-F25451-R31076"
                        /chromosome="19"
                        /map="19q13.1 between D19S208 and CAPNS"
                        /map="Overlaps CH19F14121 to the left and CH19R28052 to
                        the right"
                        /cell_type="fibroblast"
                        /map="orientation is centromere to telomere"
                        /note="Cosmid libraries constructed at LLNL from
                        flow-sorted chromosomes from hybrids U55HL9-5B and 5HL2-B,
                        which carry chromosome 19 as their only human chromosome"
repeat_region  /rpt_family="L1"
misc_feature   /rpt_family="L1"
           719..842
           /note="predicted exon, program: grail2exons_human_1.3,
           frame: 2, quality: good, score: 57.000"
repeat_region  /rpt_family="MER41"
           complement(957..1296)
repeat_region  /rpt_family="ALU"
           1289..1590
repeat_region  /rpt_family="L1"
           1907..2112
repeat_region  /rpt_family="L1"
           complement(2199..2280)
repeat_region  /rpt_family="ALU"
           complement(2310..2392)
repeat_region  /rpt_family="ALU"
           2377..2663
repeat_region  /rpt_family="ALU"
           3065..3341
repeat_region  /rpt_family="ALU"
           3429..3705
repeat_region  /rpt_family="ALU"
           <3755..4205
           /note="BLASTX similarity to (283..429): match: 0.47,
           score: 5.7e-29; database searched: nr; hypothetical L1
           protein (third intron of gene TS)- human >prf11510254A L1
           repetitive element ORF [Homo sapiens]"
repeat_region  /rpt_family="L1"
           complement(4223..4455)
repeat_region  /rpt_family="ALU"
           complement(4517..4584)
repeat_region  /rpt_family="ALU"
           complement(4822..4931)
repeat_region  /rpt_family="MIR"
           5127..5417
repeat_region  /rpt_family="ALU"
           complement(5730..6009)
repeat_region  /rpt_family="ALU"
           complement(7534..7834)
repeat_region  /rpt_family="ALU"
           complement(8629..8915)
misc_feature   /rpt_family="ALU"
           9070..9183
           /note="predicted exon, program: grail2exons_human_1.3,
           frame: 0, quality: excellent, score: 94.000"
misc_feature   10193..10276
           /note="predicted exon, program: grail2exons_human_1.3,
           frame: 1, quality: good, score: 56.000"
repeat_region  /rpt_family="ALU"
           11501..11753
repeat_region  /rpt_family="ALU"
           11801..12092
repeat_region  /rpt_family="ALU"
           complement(12570..12861)
repeat_region  /rpt_family="ALU"
           12958..13233

```

```

repeat_region  /rpt_family="ALU"
           complement(13377..13672)
repeat_region  /rpt_family="ALU"
           13899..14179
repeat_region  /rpt_family="ALU"
           14234..14310
misc_feature   /note="predicted exon, program: grail2exons_human_1.3,
           frame: 0, quality: good, score: 56.000"
           14660..14961
repeat_region  /rpt_family="ALU"
           14946..15122
repeat_region  /rpt_family="L1"
           15173..15445
repeat_region  /rpt_family="ALU"
           15584..15619
repeat_region  /rpt_family="ALU"
           complement(15626..15915)
repeat_region  /rpt_family="ALU"
           complement(15950..16258)
repeat_region  /rpt_family="ALU"
           16665..17493
repeat_region  /rpt_family="ALU"
           17507..18084
repeat_region  /rpt_family="ALU"
           complement(18269..18375)
repeat_region  /rpt_family="MER21"
           19017..19084
repeat_region  /rpt_family="ALU"
           19245..19346
repeat_region  /rpt_family="ALU"
           complement(19351..19626)
repeat_region  /rpt_family="ALU"
           19697..19833
repeat_region  /rpt_family="ALU"
           20425..20701
repeat_region  /rpt_family="ALU"
           complement(20717..20970)
misc_feature   /note="predicted exon, program: grail2exons_human_1.3,
           frame: 1, quality: marginal, score: 42.000"
           21268..21858
repeat_region  /rpt_family="ALU"
           21922..22503
repeat_region  /rpt_family="ALU"
           22690..22803
misc_feature   /note="predicted exon, program: grail2exons_human_1.3,
           frame: 0, quality: good, score: 74.000"
           complement(22986..23280)
repeat_region  /rpt_family="ALU"
           23377..23497
misc_feature   /note="predicted exon, program: grail2exons_human_1.3,
           frame: 1, quality: good, score: 57.000"
           complement(23687..23971)
repeat_region  /rpt_family="ALU"
           24001..24169
repeat_region  /rpt_family="MER21"
           24305..24583
repeat_region  /rpt_family="ALU"
           25093..25377
repeat_region  /rpt_family="ALU"
           complement(25389..25668)
repeat_region  /rpt_family="MER31"
           25459..25516
misc_feature   /note="predicted exon, program: grail2exons_human_1.3,
           frame: 0, quality: excellent, score: 79.000"
           complement(25784..26071)
repeat_region  /rpt_family="ALU"
           complement(26118..26558)
repeat_region  /rpt_family="ALU"
           26645..27013
repeat_region  /rpt_family="LTR7"
           27474..27765
repeat_region  /rpt_family="ALU"
           complement(28025..28395)

```

repeat_region /rpt_family="THE1" complement(28597..28876)
repeat_region /rpt_family="ALU" complement(29166..29667)
repeat_region /rpt_family="MER9" complement(29979..30276)
misc_feature 30918..31090 /note="predicted exon, program: grai2exons_human.1.3, frame: 1, quality: excellent, score: 95.000"
repeat_region 31106..31666 /rpt_family="ALU" complement(31848..31968)
misc_feature /note="predicted exon, program: grai2exons_human.1.3, frame: 2, quality: marginal, score: 45.000"
misc_feature complement(32692..32734) /note="predicted exon, program: grai2exons_human.1.3, frame: 2, quality: excellent, score: 94.000"
repeat_region complement(32823..33089) /rpt_family="ALU" 33296..33518
repeat_region 33296..33518 /rpt_family="MER1" 33532..33829
repeat_region /rpt_family="ALU" 33885..33983
repeat_region /rpt_family="MER1" complement(34062..34170)
misc_feature /note="predicted exon, program: grai2exons_human.1.3, frame: 1, quality: excellent, score: 97.000"
repeat_region 34376..34477 /rpt_family="ALU" 35483..35637
repeat_region /rpt_family="ALU"

Query Match 67.4%; Score 18.2; DB 10; Length 103574;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 4 CTGTGTTCTACCACTGAAGTAGG 26
||||||| || |||||
Db 22876 CTGTGTTCTCCCTCTGAACTAGG 22898

RESULT 9
CHKLKPA/c CHKLKPA 1910 bp mRNA VRT 04-FEB-1991
LOCUS Chicken cartilage link protein mRNA, complete cds.
DEFINITION M13212
ACCESSION M13212
NID 9212259
VERSION M13212.1 GI:212259
KEYWORDS
SOURCE Chicken embryo (14 day) sternal cartilage, cDNA to mRNA.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 1910)
AUTHORS Deak,F., Kiss,I., Sparks,K.J., Argaves,W.S., Hampikian,G. and Goetinck,P.F.
TITLE Complete amino acid sequence of chicken cartilage link protein deduced from cDNA clones
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 83, 3766-3770 (1986)
MEDLINE 86233315
FEATURES
source Location/Qualifiers
1..1910
/organism="Gallus gallus"
/db_xref="taxon:9031"
136..1203
/note="cartilage link protein"
/codon_start=1
/protein_id="AAA48940.1"
/db_xref="PID:q212260"
/db_xref="GI:212260"
/translation="MSSLFLVLSVCWAEPDSSLEHERIIHQENGPRLLVVA"

repeat_region /rpt_family="THE1" complement(28597..28876)
repeat_region /rpt_family="ALU" complement(29166..29667)
repeat_region /rpt_family="MER9" complement(29979..30276)
misc_feature 30918..31090 /note="predicted exon, program: grai2exons_human.1.3, frame: 1, quality: excellent, score: 95.000"
repeat_region 31106..31666 /rpt_family="ALU" complement(31848..31968)
misc_feature /note="predicted exon, program: grai2exons_human.1.3, frame: 2, quality: marginal, score: 45.000"
misc_feature complement(32692..32734) /note="predicted exon, program: grai2exons_human.1.3, frame: 2, quality: excellent, score: 94.000"
repeat_region complement(32823..33089) /rpt_family="ALU" 33296..33518
repeat_region 33296..33518 /rpt_family="MER1" 33532..33829
repeat_region /rpt_family="ALU" 33885..33983
repeat_region /rpt_family="MER1" complement(34062..34170)
misc_feature /note="predicted exon, program: grai2exons_human.1.3, frame: 1, quality: excellent, score: 97.000"
repeat_region 34376..34477 /rpt_family="ALU" 35483..35637
repeat_region /rpt_family="ALU"

EQAKIFSQRGGNVTLPCKFYHEHTSTAGSGTHKIRVKWTKLTSYLLKEVDVFVAMGHH
RKSXKYQGRVFLRESSENDASLIITNMLEDYGRKCEVTEGLEDDTAVVALNLEGV
VFYSPRLGRYNLNFHEAQACLDODSIIASFQOLYEAWRSGLDWCAAGWLSGDSVOY
PITKPRECGGKNTVPGVRNYGFWDKRSRDVFCFTSNFNGRFRYLLHPTKLTYDEA
VQACLKDAQIAKVGQIFAANKLLGYDRCDAGWLADGVSRYPIRPRKRCSPNEAAVR
FVGFPDKKHLKLYGVICFRAYN"
BASE COUNT 594 a 389 c 408 g 519 t
ORIGIN

Query Match 66.7%; Score 18; DB 4; Length 1910;
Best Local Similarity 80.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGCTTGTTCTACCACTGAAGTAGGC 27
||||||| ||| |||||
Db 276 AGCTTGTTCTGCTACCAAGTAGGC 251

RESULT 10
CHKLKPA2/c CHKLKPA2 460 bp DNA VRT 15-SEP-1990
LOCUS Chicken cartilage link protein gene, exon 3.
DEFINITION M35036
ACCESSION M35036
NID 9212262
VERSION M35036.1 GI:212262
KEYWORDS cartilage link protein.
SEGMENT 2 of 5
SOURCE Chicken (domesticus, strain White Leghorn) 9-day embryo DNA, clones lambda gup39.13 and lambda gup33.7.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 460; 49 to 63; 429 to 442)
AUTHORS Kiss,I., Deak,F., Westric,S., Delius,H., Soos,J., Dekany,K., Argaves,W.S., Sparks,K.J. and Goetinck,P.
TITLE Structure of the chicken link protein gene: Exons correlate with the protein domains
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 84, 6399-6403 (1987)
MEDLINE 87317659
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted by I. Kiss, 04-JUN-1990.
FEATURES
source Location/Qualifiers
1..460
/organism="Gallus gallus"
/db_xref="taxon:9031"
prim_transcript <1..>460
/note="cartilage link protein mRNA and introns"
intron <1..58
/note="cartilage link protein intron B"
exon 59..433
/number=3
intron 434..4460
/note="cartilage link protein intron C"
BASE COUNT 145 : 90 c 113 g 112 t
ORIGIN

Query Match 66.7%; Score 18; DB 4; Length 460;
Best Local Similarity 80.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGCTTGTTCTACCACTGAAGTAGGC 27
||||||| ||| |||||
Db 99 AGCTTGTTCTGCTACCAAGTAGGC 74

RESULT 11
NEUROIDHC 14313 bp DNA PLN 07-APR-1995
LOCUS Neurospora crassa cytoplasmic dynein heavy chain (ro-1) gene, complete cds.
DEFINITION L31504
ACCESSION L31504


```

KEYWORDS      HTG.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 139166)
Muzny,D., Arenson,A.D., Adams,C., Bunac,C., Carvelli,K., Chang,J.,
Chacko,J., Chen,J., Ding,Y., Dugan,S., Durbin,J., Forcum,J.,
Ganesh,R., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M.,
Hernandez,J., Jackson,L., Jin,S., Kampal,R., Karpathy,S., Kovar,C.,
Lau,S., Leal,B., Lee,E., Li,Y., Lichtarge,O., Liu,W., Logan,O.,
Lu,J., Ly,T., Marondel,I., Martinez,C., Merscher,S., Montgomery,K.,
Oswal,G., Perez,L., Rashid,N.D., Renault,B., Rowland,K., Savage,L.,
Scherer,S.E., Shen,H., Simon,M., Stovall,K., Timms,K.M., Todd,J.,
Vo,Q., Williamson,A., Worley,K.C., Yu,W., Kucherlapati,R.,
Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 139166)
Worley,K.C.
Direct Submission
TITLE         Submitted (27-OCT-1997) Molecular and Human Genetics, Baylor
JOURNAL       College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 139166)
Worley,K.C.
Direct Submission
TITLE         Submitted (31-JUL-1998) Human Genome Sequencing Center, Department
JOURNAL       of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
4 (bases 1 to 139166)
Worley,K.C.
Direct Submission
TITLE         Submitted (16-SEP-1998) Human Genome Sequencing Center, Department
JOURNAL       of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 139166)
Worley,K.C.
Direct Submission
TITLE         Submitted (17-SEP-1998) Human Genome Sequencing Center, Department
JOURNAL       of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 31, 1998 this sequence version replaced gi:2909690.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
FEATURES      Location/Qualifiers
source        1..139166
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="RPC13-462E2"
              /chromosome="12q24"
misc_feature  1..226
              /note="Overlaps with bases 1-226 in AC002996."
              /note="Region: Overlap with AC002996"
              join(88..201,843..945,1380..1564,6860..7024,8688..8845,
              12641..12755,18325..18759)
              /gene="Human nucleus-encoded mitochondrial aldehyde
              dehydrogenase(ALDH)gene"
              /note="Unigene cluster Hs.74630, K03001, X05409"
              530..696
repeat_region /rpt_family="AluJb"
1648..1732
/rpt_family="MIR"
complement(2001..2165)
/rpt_family="FRAM"
complement(2260..2370)
/rpt_family="L1MC3"
complement(2374..2612)
/rpt_family="AluSx"
complement(2631..2799)
/rpt_family="L1MC2"
complement(2800..3099)
/rpt_family="AluSc"
complement(3100..3220)
/rpt_family="L1MC2"
complement(3291..3588)
/rpt_family="AluJo"
complement(3608..4064)
/rpt_family="L1MB7"
complement(4069..4364)
/rpt_family="AluSx"
complement(4374..4672)
/rpt_family="AluSg"
4761..4893
/rpt_family="FLAM_C"
4904..5116
/rpt_family="AluSg/x"
5403..5684
/rpt_family="AluJb"
complement(5744..6014)
/rpt_family="AluSg"
6185..6402
/rpt_family="AluSg/x"
complement(7325..7487)
/rpt_family="FRAM"
7507..7790
/rpt_family="AluSx"
complement(7800..7943)
/rpt_family="AluSg"
complement(7947..8248)
/rpt_family="AluSx"
complement(8255..8389)
/rpt_family="FLAM_C"
complement(9202..9433)
/rpt_family="AluJo"
9452..9785
/rpt_family="L1MB6"
complement(9788..10134)
/rpt_family="THE1B"
10141..10197
/rpt_family="L1ME1"
10201..10493
/rpt_family="AluSx"
10495..10754
/rpt_family="L1MB6"
10755..10916
/rpt_family="AluSc"
10918..11213
/rpt_family="AluSg"
11214..11379
/rpt_family="AluSc"
11381..11681
/rpt_family="AluY"
11693..11988
/rpt_family="AluSg"
11996..12094
/rpt_family="(TAAA)n"
12100..12229
/rpt_family="AluJo"
12241..12529
/rpt_family="AluSx"
complement(12530..12557)
/rpt_family="AT-rich"
12631..12765
STS

```


COMMENT

Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Jul 7, 1998 this sequence version replaced gi:3228456.
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://fruitfly.berkeley.edu/sequence/>) or send
email to drosophila@hgsc.lbl.gov.
Library location: 12-74.

FEATURES

source

Location/Qualifiers
1. .51050
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/map="22E1-22E1"
/clone="P1 DS07020 (D238)"
14368 a 11189 c 11359 g 14134 t
BASE COUNT 14368 a 11189 c 11359 g 14134 t
ORIGIN

Query Match 66.7%; Score 18; DB 36; Length 51050;
Best Local Similarity 80.8%; Pred. NO. 1.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 2 AGCTTGTTCTACCACTGAAGTGGC 27
||||| || ||||| ||||| |||||
Db 9746 AACTGTGTCACCACTGAAGTGGC 9771

RESULT 15

AC001657/c

LOCUS

DEFINITION Drosophila melanogaster DNA sequence (P1 DS06874 (D19)), complete
sequence.
AC001657 L36286 L36305 L36304 L36289 L36300 L39666 L36301 L36298
L81427 L36303 L81426 L39670 L39668 L36294 L36285 L36295 L39667
L36296 L36287 L36299 L36291 L36297 L36284 L39671 L39665 L39669
L36292 L36293 L36302 L36290
92337899

NID

AC001657.1 GI:2337899

VERSION

HTG.

KEYWORDS

SOURCE Drosophila melanogaster (Subclones in sac from P1 clone DS06874

ORGANISM

(D19)) DNA.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE

AUTHORS

1 (bases 1 to 85139)
Celisner, S.E., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E.,
Svirskas, R.R., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,
Blazej, R.G., Housten, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
Flanagan, J., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S.,
Kim, S.H., Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
Moshrefi, A.R., Panch, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R.,
Yee, A., Zhang, R., Zieran, L.L. and Kimmel, B.
Sequencing of Alcohol dehydrogenase region

TITLE

JOURNAL

REFERENCE

AUTHORS

Unpublished (1997)
2 (bases 1 to 85139)
Martin, C.H., Arcaina, T.T., Bondoc, M.M., Chiang, A., Critz, P.A.,
Davis, C.A., Doyle, C.M., Ericsson, C.L., Farfan, D.E., Gunning, K.M.,
Houston, K.A., Jaklevic, M.A., Kadner, K.E., Kim, K., Kim, S.F.,
Ko, C.L., Lewis, K.D., Li, M., Lindquist, K.J., Lomotan, M.A.,
Lustre, V.M., Machrus, M.U., Mayeda, C.A., Miguel, T.M., Miller, C.A.,
Mok, M.S., Pacleb, J.M., Patel, S.G., Santos, R.F., Subramanian, S.,
Wan, K.H., Whitelaw, K.R., Yee, A., Yeh, R.T., Yu, C. and Palazzolo, M.J.
Direct Submission

TITLE

JOURNAL

COMMENT

Submitted (22-APR-1997)
On Aug 20, 1997 this sequence version replaced gi:1945585.
Sequence submitted by:

Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720

For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive Web site (<http://fruitfly.berkeley.edu/sequence/>) or send
email to drosophila@hgsc.lbl.gov.
Library location: 58-72.

FEATURES

source

Location/Qualifiers
1. .85139
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/map="3B7-35C1"
/clone="P1 DS06874 (D19)"
/note="This sequence has not changed since its original
submission on 04/22/1997. It was resubmitted in order to
include all secondary accession numbers for the subclones
belonging to this clone."
BASE COUNT 25995 a 16491 c 16856 g 25797 t
ORIGIN

Query Match 66.7%; Score 18; DB 37; Length 85139;
Best Local Similarity 80.8%; Pred. NO. 1.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 AAGCTTGTTCTACCACTGAAGTGG 26
||||| || ||||| ||||| |||||
Db 41560 AAGCTGTTTATACCATCTGACCAGG 41535

Search completed: September 18, 1999, 15:49:42
Job time: 1462 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 05:27:26 ; Search time 213.04 Seconds
(without alignments)
31.709 Million cell updates/sec

Title: US-09-037-472-9
Perfect score: 27
Sequence: 1 AAGCTTGTTTACCACTGAAGTAGGC 27

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	27	100.0	27	V32389	Interleukin-1-alpha
2	22	81.5	22	T70316	Primer for detecti
3	22	81.5	5701	X02988	Human IL-1ra BAC c
4	21	77.8	21	X16611	Interleukin 1 (441
5	18	66.7	327	T78683	Staphylococcus aur
6	17.4	64.4	633	T75647	Staphylococcus aur
7	17.2	63.7	687	T92382	Bacillus smithii n
8	17.2	63.7	648	T77989	Staphylococcus aur
9	17	63.0	1253	N40312	Sequence of human
10	17	63.0	6422	T20576	Polynucleotide seq
11	16.8	62.2	8160	X13096	Enterococcus faeca
12	16.6	61.5	7753	Q26727	Encodes exons XV t
13	16.6	61.5	59	T79270	Staphylococcus aur
14	16.6	61.5	400	T78290	Staphylococcus aur
15	16.6	61.5	361	T78258	Staphylococcus aur
16	16.6	61.5	400	T78053	Staphylococcus aur
17	16.6	61.5	400	T78035	Staphylococcus aur
18	16.6	61.5	401	T78042	Staphylococcus aur
19	16.6	61.5	553	T78001	Staphylococcus aur
20	16.6	61.5	400	T77971	Staphylococcus aur
21	16.6	61.5	400	T77942	Staphylococcus aur
22	16.6	61.5	475	T77932	Staphylococcus aur
23	16.6	61.5	1171	T77850	Staphylococcus aur
24	16.6	61.5	400	T77853	Staphylococcus aur
25	16.6	61.5	6591	T77425	Staphylococcus aur
26	16.6	61.5	237	T78877	Staphylococcus aur
27	16.6	61.5	239	T78855	Staphylococcus aur
28	16.6	61.5	239	T78854	Staphylococcus aur
29	16.6	61.5	411	T78654	Staphylococcus aur
30	16.6	61.5	339	T78574	Staphylococcus aur
31	16.6	61.5	400	T78545	Staphylococcus aur
32	16.6	61.5	381	T78334	Staphylococcus aur
33	16.6	61.5	400	T78340	Staphylococcus aur
34	16.6	61.5	1421	X13384	Enterococcus faeca
35	16.4	60.7	1393	T36747	Sequence of TB-1 w
36	16.4	60.7	2259	T46782	40 kDa and 34 kDa
37	16.2	60.0	384	T21919	Human gene signatu
38	16.2	60.0	1557	T70441	H. pylori AlpB gen
39	16.2	60.0	1557	T70248	DNA encoding Helic
40	16.2	60.0	4190	T34854	Human retinal dege
41	16.2	60.0	1922	T84426	Kidney injury asso
42	16.2	60.0	1838	T80605	
43	16.2	60.0	110000	X20248_03	Continuation (4 of

44 16.2 60.0 5504 1 X20559 Polynucleotide seq
c 45 16.2 60.0 73 1 X30293 Borrelia burgdorfe

ALIGNMENTS

RESULT 1

V32389
ID V32389 standard; DNA; 27 BP.
AC V32389, 1998 (first entry)
DT Interleukin-1-alpha primer 1 (-967/-945).
DE IL-1-alpha; genetic polymorphism; PCR; primer; amplification;
KW sight threatening diabetic retinopathy; interleukin-1-alpha;
KW interleukin-1-beta; interleukin-1RN; ss.
OS Synthetic.
OS Homo sapiens.
PN W09815653-A1.
PD 16-APR-1998.
PF 09-OCT-1997; G02790.
PR 10-OCT-1996; GB-021129.
PA (DUFF/) DUFF G.
PA (RENN/) RENNIE I.
PA (RICH/) RICHARDSON R.
PI Duff G, Rennie I, Richardson R;
WPI: 98-240835/21.
DR Predicting increased risk of sight-threatening diabetic retinopathy
PT - comprises identifying genetic polymorphism pattern for genes
IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
symptoms occur
PS Claim 2: Page 33: 41pp; English.
CC Interleukin-1-alpha (IL-1-alpha) primers 1 and 2 (V32390) were used to
amplify the IL-1-alpha gene region to identify single base variation
polymorphism of C/T at base 889. The invention claims to provide a
method for predicting the risk of sight threatening diabetic retinopathy.
CC The method involves isolating DNA from a patient and determining the DNA
polymorphism pattern of the genes that code for interleukin-1-alpha,
interleukin-1-beta and interleukin-1RN. The polymorphic pattern
identified is then compared with controls of known DNA polymorphism
patterns thereby identifying patients carrying a genetic polymorphism
associated with increased risk of sight threatening diabetic retinopathy.
CC The method may be able to identify diabetic patients at risk before the
clinically detectable disorders occur. Polymorphism pattern
determination of IL genes involved PCR reactions using primers V32389-
V32390. The method is also claimed to be useful in conjunction with
identification of other genes associated with sight threatening diabetic
retinopathy in genomic DNA and therefore, in identifying diabetic
patients expressing multiple risk patterns. 7 A; 8 C; 5 G; 7 T;
SQ Sequence 27 BP; 7 A; 8 C; 5 G; 7 T;

Query Match 100.0%; Score 27; DB 1; Length 27;

Best Local Similarity 100.0%; Pred. No. 0.00056;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AAGCTTGTTTACCACTGAAGTAGGC 27

Db 1 AAGCTTGTTTACCACTGAAGTAGGC 27

RESULT 2

T70316
ID T70316 standard; DNA; 22 BP.
AC T70316;
DT 19-OCT-1997 (first entry)
DE Primer for detecting genetic predisposition to periodontal disease.
KW Periodontal disease; gingivitis; periodontitis; polymorphism;
KW interleukin-1 alpha; IL-1A; interleukin-1 beta; IL-1B; primer; PCR;
KW polymerase chain reaction; ss.
OS Synthetic.
PN W09706180-A1.
PD 20-FEB-1997.

```
PF 02-AUG-1996; U12455.
PR 03-AUG-1995; US-510696.
PA (KORN/) KORNMAN K S.
PA (MEDI-) MEDICAL SCI SYSTEMS INC.
PI Duff GW, Kornman KS;
DR WPI: 97-154207/14.
PT Identification of patient's genetic polymorphism pattern - allows
PT Prediction of increased periodontal disease severity
PS Claim 3; Page 25; 33pp; English.
CC PCR primers (T70316 and T13884-88) can be used in a new method
CC for identifying a genetic predisposition to periodontal disease by
CC detecting the presence of DNA polymorphisms in the gene sequences
CC for interleukin-1 alpha (IL-1a) and interleukin-1 beta (IL-1b).
CC Alleles associated with severe disease were identified as IL-1a
CC allele 2 together with IL-1b (Taql) allele 2. The primers are used
CC to amplify DNA from a blood or tissue sample and products are
CC subjected to restriction digestion to determine the polymorphism
CC pattern of the patient. A single base variation (C/T) polymorphism
CC at IL-1a base -889 can be identified using primers (T70316 and
CC T13884) corresponding to bases -967 to -945 and -888 to -869,
CC respectively. An NcoI site is created if C is available at -889,
CC but not if T is present.
SQ Sequence 22 BP; 5 A; 7 C; 4 G; 6 T;

Query Match 81.5%; Score 22; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TGTCTACCACTGAAGTAGGC 27
DB 1 TGTCTACCACTGAAGTAGGC 22

RESULT 3
X02988
ID X02988 standard; DNA: 5701 BP.
AC X02988;
DT 22-JUN-1999 (first entry)
DE Human IL-1ra BAC contiguous DNA sequence 33.
KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammation; bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
PN W0906436-A1.
PD 11-FEB-1998.
PF 03-AUG-1998; U16102.
PR 02-JUL-1998; US-091650.
PR 04-AUG-1997; US-054646.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Pan Y;
DR WPI: 99-153692/13.
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
PS Example 5; Figure 3; 22pp; English.
CC X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC containing alternatively spliced forms of human IL-1ra. Such fragments
CC are used in the method of the invention which describes the isolation of
CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC member of the cytokine superfamily that is expected to inhibit
CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC also bind to a new receptor so could regulate other cellular processes
CC associated with acute or chronic inflammation, e.g. asthma, chronic
CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent
CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 5701 BP; 1729 A; 1238 C; 1079 G; 1654 T;

Query Match 81.5%; Score 22; DB 1; Length 5701;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TGTCTACCACTGAAGTAGGC 27
DB 1 TGTCTACCACTGAAGTAGGC 22

RESULT 4
X16611
ID X16611 standard; DNA: 21 BP.
AC X16611;
DT 29-APR-1999 (first entry)
DE Interleukin 1 (44112332) haplotype PCR primer #5.
KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;
KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;
KW Graves disease; systemic lupus erythematosus; lichen sclerosis;
KW ulcerative colitis; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN W09854359-A1.
PD 03-DEC-1998.
PF 21-MAY-1998; G01481.
PR 29-MAY-1997; GB-011040.
PA (CAMP/) CAMP N J.
PA (COXA/) COX A.
PA (DUFF/) DUFF G, De Giovine FS, Duff G;
PI Camp NJ, Cox A, De Giovine FS, Duff G;
DR WPI: 99-080814/07.
PT New method of determining a patient's susceptibility to inflammatory
PT disorders - by detecting the presence of an IL-1 (44112332)
PT haplotype, useful in designing treatment strategies that modulate
PT the activity of proteins produced by the IL-1 gene cluster
PS Claim 3; Page 33; 49pp; English.
CC A method has been developed for determining a patient's susceptibility
CC to an inflammatory disorder. The method comprises the detection of an
CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the
CC patient, where its presence indicates susceptibility to an inflammatory
CC disorder. X16607 to X16631 represent PCR primers used in the method for
CC detecting the IL-1 (44112332) haplotype. The method provides kits for
CC the early prediction of a patient's susceptibility to inflammatory
CC disorders, including coronary artery disease, osteoporosis, nephropathy
CC in diabetes mellitus, alopecia areata, Graves disease, systemic lupus
CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of
CC alleles of the haplotype can be applied to particular inflammatory
CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,
CC psoriasis, and insulin dependent diabetes. The identification of a
CC disease-associated haplotype enables determination of which alleles are
CC causative, and this information is useful in designing treatment
CC strategies, including gene therapy and treatment using various agents
CC that modulate the activity of proteins produced by the IL-1 gene cluster.
CC Some alleles from the IL-1 gene cluster are associated with particular
CC inflammatory diseases, and insufficient IL-1 production appears to act
CC centrally in the pathology of these diseases. Therefore, the use of IL-1
CC gene clusters is useful in determining genetic susceptibility to
CC inflammatory diseases, including those with a multifactorial etiology
CC with a polygenic component.
SQ Sequence 21 BP; 5 A; 6 C; 4 G; 6 T;

Query Match 77.8%; Score 21; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.33;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TGTCTACCACTGAAGTAGGC 26
DB 1 TGTCTACCACTGAAGTAGGC 21

RESULT 5
V78683
ID V78683 standard; DNA: 327 BP.
```

V78683;
 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #4372.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 PN EP-786519-A2.
 PD 30-JUL-1997.
 PF 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
 stored on computer readable medium and used in the production of
 anti-S.aureus vaccines
 Claim 1; Page 2338; 3271pp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 of the invention. The DNA sequences are recorded on a computer readable
 medium, preferably selected from a floppy or hard disk, random access
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 the S.aureus DNA sequences allows putative functions to be assigned so
 that protein-encoding or regulatory regions of commercial, therapeutic or
 industrial importance can be obtained. Specifically, sequences which are
 likely to encode antigens have been identified and these polypeptides can
 be used in a vaccine composition against S.aureus infection. The
 polypeptides can also be used in a kit for the immunodetection of
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 skin and surgical wound infections, scalded skin syndrome, toxic shock
 syndrome, etc. Organisms transformed with the DNA sequences can be used
 for recombinant production of the polypeptides. The new DNA sequences
 (and their fragments) are useful as primers or probes for isolating
 homologues of any of the S.aureus DNA sequences contained on the
 computer readable medium.
 Sequence 327 BP; 87 A; 85 C; 75 G; 77 T;
 Query Match 66.7%; Score 18; DB 1; Length 327;
 Best Local Similarity 80.8%; Pred. No. 12;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 AGCTTGTCTACCACTGAACCTAGGC 27
 ||||| ||||| ||||| |||||
 Db 170 AGTTGTCTACCAATTGACCTAGGC 195
 RESULT 6
 V75647/c
 ID V75647 standard; DNA; 633 BP.
 AC V75647;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #1336.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 PN EP-786519-A2.
 PD 30-JUL-1997.
 PF 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
 stored on computer readable medium and used in the production of
 anti-S.aureus vaccines
 Claim 1; Page 1924; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 of the invention. The DNA sequences are recorded on a computer readable
 medium, preferably selected from a floppy or hard disk, random access
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 the S.aureus DNA sequences allows putative functions to be assigned so
 that protein-encoding or regulatory regions of commercial, therapeutic or
 industrial importance can be obtained. Specifically, sequences which are
 likely to encode antigens have been identified and these polypeptides can
 be used in a vaccine composition against S.aureus infection. The
 polypeptides can also be used in a kit for the immunodetection of
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 skin and surgical wound infections, scalded skin syndrome, toxic shock
 syndrome, etc. Organisms transformed with the DNA sequences can be used
 for recombinant production of the polypeptides. The new DNA sequences
 (and their fragments) are useful as primers or probes for isolating
 homologues of any of the S.aureus DNA sequences contained on the
 computer readable medium.
 Sequence 633 BP; 252 A; 105 C; 90 G; 184 T;
 Query Match 64.4%; Score 17.4; DB 1; Length 633;
 Best Local Similarity 94.7%; Pred. No. 25;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 5 TTGTTCTACCACTGAACCT 23
 ||||| ||||| ||||| |||||
 Db 86 TTGTCCTACCACTGAACCT 68

RESULT 7
 T92382/c
 ID T92382 standard; DNA; 687 BP.
 AC T92382;
 DT 28-JAN-1998 (first entry)
 DE Bacillus smithii nitrile hydratase subunit beta encoding DNA.
 KW Nitrile hydratase subunit alpha; nitrile hydratase subunit beta;
 KW acrylonitrile; acryloamide; biological catalysis; amide;
 KW thermally stable protein; ds.
 OS Bacillus smithii.
 FH Key Location/Qualifiers
 FT CDS 1..687
 FT /*tag= a
 FT /product= Nitrile_hydratase_subunit_beta
 FT /note= "No stop codon given"
 PN J09248188-A.
 PD 22-SEP-1997.
 PF 18-MAR-1996; 060732.
 PR 18-MAR-1996; JP-060732.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI: 97-520742/48.
 DR P-PSDB; W32621.
 PT Gene encoding nitrile hydratase - for producing amide from nitrile
 by biological catalysis
 PS Claim 7; Page 11-12; 16pp; Japanese.
 CC The present sequence encodes nitrile hydratase subunit beta, a novel
 CC protein isolated from Bacillus smithii. The protein has hydration
 CC activity for converting acrylonitrile into acryloamide. It is useful
 CC for producing amide from nitrile by biological catalysis.
 SQ Sequence 687 BP; 212 A; 112 C; 173 G; 190 T;
 Query Match 63.7%; Score 17.2; DB 1; Length 687;
 Best Local Similarity 86.4%; Pred. No. 32;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 3 GCCTTGTCTACCACTGAACCTA 24
 ||||| ||||| ||||| |||||
 Db 374 GCCTTGTCTACCACTGAACCTA 353
 RESULT 8
 V77989

ID V77989 standard; DNA; 648 BP.
AC V77989;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #3678.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 121..180 /*tag= a
FT /*note= *these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence*
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon FJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
DT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 2639-2640; 327lpp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 648 BP; 161 A; 150 C; 144 G; 131 T;
Query Match 63.7%; Score 17.2; DB 1; Length 648;
Best Local Similarity 82.6%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 5 TTGTTCTACCACTGAACTAGGC 27
DB 9 TTGNTCTACCAATGAGCTAGGC 31
RESULT 9
N40312
ID N40312 standard; DNA; 1253 BP.
AC N40312;
DT 03-AUG-1992 (first entry)
DE Sequence of human leukocyte interferon (leIFN) gene.
KW Interferon; yeast expression vector; promoter; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 52..621 /*tag= a
FT W08401153-A.
PN 29-MAR-1984.
PD 09-SEP-1983; U01370.
PF 15-SEP-1982; US-418521.

PA (COLB) COLLABORATIVE RES INC.
PI Rotstein D, Bowden DW, Davis RW, Fink GR, Knowlton RG,
PI Mao J, Taunton-Rigby A, Vovis GF;
DR WPI: 84-088509/14.
PS P-PSDB: P40741.
PT Prodn. of interferon in yeast - by use of recombinant DNA segment
PT contg. suc 2 promoter linked to interferon gene
PS Example; Page 48; 53pp; English.
CC The inventors claim a method for the prodn. of interferon in yeast
CC by use of recombinant DNA segment contg. suc 2 promoter linked to
CC interferon gene. The interferon synthesis can be regulated over a
CC broad range by changes in the culture medium, e.g. synthesis is
CC repressed in glucose medium and de-repressed when glucose is
CC depleted. Also claimed is yeast strain ATCC 20644-CGY144.
SQ Sequence 1253 BP; 373 A; 268 C; 244 G; 368 T;
Query Match 63.0%; Score 17; DB 1; Length 1253;
Best Local Similarity 80.0%; Pred. No. 42;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1 AAGCTTGTTCTACCACTGAACTAG 25
DB 2 AAGCTTGCTCATCCATCTGAACCAG 26
RESULT 10
X20576
ID X20576 standard; DNA; 6422 BP.
AC X20576;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN W08859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
PI WPI: 99-081273/07.
DT New isolated Treponema pallidum nucleic acids - used to develop
DT products for the detection, diagnosis, characterisation, prevention
DT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1; Page 593-597; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 6422 BP; 1614 A; 1888 C; 1547 G; 1361 T;
Query Match 63.0%; Score 17; DB 1; Length 6422;
Best Local Similarity 80.0%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 3 GCTTGTCTTACCACTGAACTAGGC 27
DB 1198 GCGTGCTCTACCACTGAGCTATGC 1222
RESULT 11
X13096
ID X13096 standard; DNA; 8160 BP.
AC X13096;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:159.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.

PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI: 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 898-902; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 8160 BP; 2731 A; 1463 C; 1654 G; 2292 T;
Query Match 62.2%; Score 16.8; DB 1; Length 8160;
Best Local Similarity 90.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 6 TGTCTACCACTGAACTAG 25
DB 3366 TGTCTAGCTCCTGAACTAG 3385
RESULT 12
ID Q26727/c
AC Q26727 standard; DNA; 7753 BP.
DT 20-JAN-1993 (first entry)
DE Encodes exons XV to XVIII of human hepatocyte growth factor.
KW HGF; enhance growth; preparing transgenic animals; hepatic disease;
KW clinical diagnostic reagent; drug.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 101..241
FT /*tag= a
FT /*note= "exon XV"
FT cds 773..879
FT /*tag= b
FT /*note= "exon XVI"
FT cds 991..1136
FT /*tag= c
FT /*note= "exon XVII"
FT cds 3764..3937
FT /*tag= d
FT /*note= "exon XVIII"
FT polya_signal 7492..7497
FT /*tag= e
PN J04183394-A.
PD 30-JUN-1992.
PF 19-NOV-1990; 314548.
PR 19-NOV-1990; JP-314548.
PA (NAKA/) NAKAMURA T.
PA (TOYK) TOYOBO KK.
DR WPI: 92-265591/32.
DR P-PSDB; R25692.
PT Recombinant human hepatocyte growth factor and DNA encoding it -

PT useful for dia nosis and treatment of hepatic disease and
PT transgenic animal prepn.
PS Disclosure: Page 22; 28pp; Japanese.
CC This sequence contains exons XV to XVIII of human hepatocyte growth
CC factor. See also R25676-92, Q26713-27.
SQ Sequence 7753 BP; 2567 A; 1177 C; 1243 G; 2766 T;
Query Match 61.5%; Score 16.6; DB 1; Length 7753;
Best Local Similarity 82.6%; Pred. No. 84;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 AGCTTGTTCTACCACTGAACTA 24
DB 6273 ACCCTGTTCTACCACTGAACTA 6251
RESULT 13
ID V79270
AC V79270 standard; DNA; 59 BP.
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #4959.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 3132; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 59 BP; 16 A; 18 C; 11 G; 14 T;
Query Match 61.5%; Score 16.6; DB 1; Length 59;
Best Local Similarity 82.6%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 5 TTGTTCTACCACTGAACTAGGC 27
DB 35 TTGCTCTACCAATTGAGCTAGGC 57
RESULT 14
ID V78290
DT V78290 standard; DNA; 400 BP.

AC V78290;
 DE 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #3979.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 FH Key Location/Qualifiers
 FT misc_feature 121..180
 FT /*tag= a
 FT /note= "these bases represent a line of missing text in
 the sequence listing in the specification. They
 are included to maintain the nucleotide numbering
 given in the specification. for this DNA sequence"
 FT
 PN EP-786519-A2.
 PD 30-JUL-1997.
 PF 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI: 97-374922/35.
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1: Page 2772-2773: 3271pp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 400 BP; 87 A; 90 C; 84 G; 75 T;
 Query Match 61.5%; Score 16.6; DB 1; Length 400;
 Best Local Similarity 82.6%; Pred. No. 56;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 5 TTGTTCTACCACTGAGCTAGGC 27
 Db 294 TTGCTCTACCAATGAGCTAGGC 316
 RESULT 15
 V78258
 ID V78258 standard; DNA; 361 BP.
 AC V78258.
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #3947.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 PN EP-786519-A2.
 PD 30-JUL-1997.
 PF 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI: 97-374922/35.
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1: Page 2759: 3271pp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 361 BP; 104 A; 83 C; 78 G; 93 T;
 Query Match 61.5%; Score 16.6; DB 1; Length 361;
 Best Local Similarity 82.6%; Pred. No. 55;
 Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
 QY 5 TTGTTCTACCACTGAGCTAGGC 27
 Db 47 TTGCTCTACCAATGAGCTAGGC 69

Search completed: September 18, 1999, 05:27:26
 Job time: 1691 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:49 ; Search time 1405 Seconds
(without alignments)
37.906 Million cell updates/sec

Title: US-09-037-472-9

Perfect score: 27
Sequence: 1 AAGCTTGTTCTACCCACCTGAACCTAGGC 27

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.6	68.9	405	25	N88530	N88530 K3598F Huma
C 2	18.6	68.9	525	36	AA607268	AA607268 vm92h05.i
C 3	18.6	68.9	459	37	AA681861	AA681861 v44a05.s
C 4	18	66.7	224	32	AA331881	AA331881 EST335751
C 5	18	66.7	564	33	AA422885	AA422885 v447c06.s
C 6	18	66.7	422	39	AA824084	AA824084 v166c03.s
C 7	17.8	65.9	465	43	AI161663	AI161663 A004P75U
C 8	17.6	65.2	337	20	T05736	T05736 ESTQ3625 Fe
C 9	17.6	65.2	308	21	F13213	F13213 HSC3K111 n
C 10	17.6	65.2	526	21	T60600	T60600 yb91g05.rl
C 11	17.6	65.2	425	27	AA004244	AA004244 zh88d06.i
C 12	17.6	65.2	480	27	AA004527	AA004527 zh91a09.i
C 13	17.6	65.2	593	48	AI566060	AI566060 tn52h04.x
C 14	17.6	65.2	226	50	AV046966	AV046966 AV046966
C 15	17.4	64.4	399	20	T10245	T10245 seq1305 b4H
C 16	17.4	64.4	308	20	T32622	T32622 EST52100 Hu
C 17	17.4	64.4	294	20	T32523	T32523 EST81965 Hu
C 18	17.4	64.4	267	20	T35508	T35508 EST86615 Hu
C 19	17.4	64.4	323	20	T35551	T35551 EST87186 Hu
C 20	17.4	64.4	281	20	T35556	T35556 EST87349 Hu
C 21	17.4	64.4	425	21	R17871	R17871 Y910d02.rl
C 22	17.4	64.4	383	21	R18749	R18749 Y917c08.rl
C 23	17.4	64.4	484	22	R54291	R54291 Y978f06.rl
C 24	17.4	64.4	442	24	H94074	H94074 YW58a10.rl
C 25	17.4	64.4	273	31	AA300763	AA300763 EST13619
C 26	17.4	64.4	556	31	AA306372	AA306372 EST177333
C 27	17.4	64.4	549	31	AA306551	AA306551 EST177501
C 28	17.4	64.4	395	31	AA307033	AA307033 EST177948
C 29	17.4	64.4	503	31	AA307495	AA307495 EST178379
C 30	17.4	64.4	594	31	AA308049	AA308049 EST178890
C 31	17.4	64.4	571	31	AA310973	AA310973 EST181747
C 32	17.4	64.4	528	31	AA311686	AA311686 EST182410
C 33	17.4	64.4	401	31	AA314612	AA314612 EST186444
C 34	17.4	64.4	354	31	AA316439	AA316439 EST188137
C 35	17.4	64.4	407	47	AI508192	AI508192 mh31e12.y
C 36	17.4	64.4	815	47	AI525561	AI525561 PT1.3.03
C 37	17.4	64.4	670	49	C96306	C96306 C96306 Marc
C 38	17.4	64.4	459	49	C96449	C96449 C96449 Marc
C 39	17.4	64.4	708	53	HSN001010	HSN001010 Homo sapi
C 40	17.2	63.7	280	20	D39825	D39825 R1C51442A R
C 41	17.2	63.7	452	38	AA808387	AA808387 oa90d09.s
C 42	17.2	63.7	443	39	AA860156	AA860156 ak47b09.s
C 43	17.2	63.7	238	48	AI558929	AI558929 fb78b02.y
C 44	17	63.0	334	23	HS1722	HS1722 YP85e07.rl
C 45	17	63.0	502	54	HSN010701	HSN010701 Homo sapi

ALIGNMENTS

RESULT 1
N88530/c
LOCUS N88530 405 bp mRNA EST 02-APR-1996
DEFINITION K3598F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA
ACCESSION clone K3598 5', mRNA sequence.
NID N88530
VERSION q1441732
N88530.1 GI:1441732


```

/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone_lib="IMAGE:1123472"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pBluescribe (modified);
Site:1: MluI; Site:2: SalI; Cloned unidirectionally from
mRNA prepared from 13,500 2-cell stage embryos. Primer:
SalI(dT): 5'-CGGTGACGTCGACCGTTTATTTT-3'.
CDNAs
were cloned into the MluI/SalI sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."
BASE COUNT      105 a   112 c   101 g   141 t
ORIGIN
Query Match      68.9%; Score 18.6; DB 37; length 459;
Best Local Similarity 84.0%; Pred. No. 54;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTGTGTTACACCTGAACCTAG 25
|||||  |||||  |||||  |||||  |||||
Db 167 AAGCTGTGTTGCCACCTCACCCTAG 143

RESULT 4
AA331881/c
LOCUS
DEFINITION
EST35751 Embryo, 8 week I Homo sapiens cDNA 5' end similar to
similar to MER5 protein, mRNA sequence.
ACCESSION
AA331881
NID
g1984123
VERSION
AA331881.1 GI:1984123
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 224)
AUTHORS
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man,Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S., Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Snail,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Bednarek,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseltine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
Nature 377 (6547 Suppl), 3-174 (1995)
96026280
On Sep 12, 1996 this sequence version replaced gi:1405042.
Other_ESTs: THC170844
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018689056
Fax: 3018689423
Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene

```

```

Index (http://www.tigr.org/tldb/hgi/hgi.html)
Seq primer: M13 Reverse.

```

FEATURES

Source

```

1..224
/organism="Homo sapiens"
/db_xref="Arcc (inhost):133506"
/db_xref="taxon:9606"
/clone_lib="Embryo, 8 week I"
/dev_stage="embryo, 8 wks"
/notes="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
Site:1: EcoRI; Site:2: XhoI"

```

BASE COUNT

```

42 a   60 c   60 g   56 t
ORIGIN

```

```

Query Match      66.7%; Score 18; DB 32; Length 224;
Best Local Similarity 77.8%; Pred. No. 88;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

QY 1 AAGCTTGTGTTACACCTGAACCTAGGC 27
|||||  |||||  |||||  |||||  |||||

```

```

Db 149 AAGCTGCTGTTCCACATGCAGNAGGC 123

```

RESULT

```

AA422885
LOCUS
DEFINITION

```

```

AA422885      564 bp      mRNA      EST      16-OCT-1997
vd47c06.sl Knowles Solter mouse 2 cell Mus musculus cDNA clone
IMAGE:803722 5' similar to gb:X64550 M.musculus mRNA for
hyaluronan-mediated motility (MOUSE);, mRNA sequence.
ACCESSION
AA422885
NID
92101727
VERSION
AA422885.1 GI:2101727
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 564)
AUTHORS
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On May 8, 1995 this sequence version replaced gi:800408.

```

TITLE

```

JOURNAL
COMMENT

```

```

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:480066
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 441.

```

FEATURES

source

```

1..564
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/map="15"
/clone="IMAGE:803722"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pBluescribe (modified);
Site:1: MluI; Site:2: SalI; Cloned unidirectionally from

```

mRNA prepared from 13,500 2-cell stage embryos. Primer:
Sali(dfr): 5'-CGGTGACCGTCGACGTTTTTTTTTTT-3'. cDNAs
were cloned into the MluI/Sali sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 kb."

BASE COUNT 159 a 97 c 176 g 132 t
ORIGIN

Query Match 66.7%; Score 18; DB 33; Length 564;

Best Local Similarity 80.8%; Pred. No. 1.1e+02;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGCTGTGTTCTACCACTGAAGTGGC 27

I | | | | | | | | | | | | | | | | | | | | |

Db 176 ATCTGTGTTCTGCGAGCTGAAGTGGC 201

RESULT 6

AA824084

LOCUS AA824084 422 bp mRNA EST 17-FEB-1998

DEFINITION vr66c03.s1 Knowles Solter mouse 2 cell Mus musculus cDNA clone

IMAGE:1125604 5' similar to gb:X64550 M.musculus mRNA for

hyaluronan-mediated motility (MOUSE);, mRNA sequence.

ACCESSION AA824084

NID 92893952

VERSION AA824084.1 GI:2893952

KEYWORDS EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 422)

AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,

Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,

Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,

Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and

Waterston,R.

TITLE The WashU-HMNI Mouse EST Project

JOURNAL Unpublished (1996)

COMMENT On Sep 12, 1996 this sequence version replaced gi:1404594.

Contact: Marra M/Mouse EST Project

WashU-HMNI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:614940

High quality sequence stop: 422.

Location/Qualifiers

1..422

/organism="Mus musculus"

/strain="B6D2 F1/J"

/db_xref="taxon:10090"

/clone="IMAGE:1125604"

/clone_lib="Knowles Solter mouse 2 cell"

/tissue_type="embryo"

/dev_stage="2-cell"

/lab_host="DH10B"

/note="Organ: embryo; Vector: pBluescribe (modified);

Site 1: MluI; Site 2: Sali; Cloned unidirectionally from

mRNA prepared from 13,500 2-cell stage embryos. Primer:

Sali(dfr): 5'-CGGTGACCGTCGACGTTTTTTTTTTT-3'. cDNAs

were cloned into the MluI/Sali sites of a modified

pBluescribe vector using commercial linkers (NEB).

Average insert size: 1.2 kb."

BASE COUNT 116 a 70 c 130 g 106 t

ORIGIN

Query Match 66.7%; Score 18; DB 39; Length 422;

Best Local Similarity 80.8%; Pred. No. 1.1e+02;

Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 2 AGCTGTGTTCTACCACTGAAGTGGC 27

I | | | | | | | | | | | | | | | | | | | | |

Db 182 ATCTGTGTTCTGCGAGCTGAAGTGGC 207

RESULT 7

AI161663/c

LOCUS AI161663 465 bp mRNA EST 03-DEC-1998

DEFINITION A004P75U Hybrid aspen plasmid library Populus tremula x Populus

tremuloides cDNA 5', mRNA sequence.

ACCESSION AI161663

NID 93852948

VERSION AI161663.1 GI:3852948

KEYWORDS EST.

SOURCE Populus tremula x Populus tremuloides.

ORGANISM Populus tremula x Populus tremuloides

Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core

eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Populus.

1 (bases 1 to 465)

AUTHORS Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Riede,A.,

Holmberg,A., Amini,B., Bhaleao,R., Larsson,M., Villarroel,R., Van

Montagu,M., Sandberg,G., Olsson,O., Teeri,T.T., Boerjan,W.,

Gustafsson,P., Uhlen,M., Sundberg,B. and Lundberg,J.

Gene discovery in the wood-forming tissues of poplar: Analysis of

5,692 expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)

99007314

COMMENT On Jan 9, 1998 this sequence version replaced gi:930585.

Contact: Sterky F

Department of Biotechnology

Royal Institute of Technology (KTH)

Teknikringen 34, S-100 44 STOCKHOLM, Sweden

Tel: +46 8 790 8287

Fax: +46 8 24 54 52

Email: fredrik@biochem.kth.se

PCR Primers

FORWARD: AAAGGGGATGTGCTGCAAGGCG

BACKWARD: GCTTCGGCTGCTATGTTGCTG

Seq primer: CGTTGAAACGACGCCACG

High quality sequence stop: 465.

Location/Qualifiers

1..465

/organism="Populus tremula x Populus tremuloides"

/db_xref="taxon:47664"

/clone_lib="Hybrid aspen plasmid library"

/tissue_type="Cambial region"

/dev_stage="1.5 m actively growing tree"

/lab_host="E.coli"

/note="Vector: pBluescript SK; Site 1: Sali; Site 2: NotI;

Cambial region tissues, including developing xylem, the

meristematic cambial zone and the developing and mature

phloem, was harvested from 1.5 m actively growing trees.

cDNA was prepared and cloned into lambda gt2a. DNA was

isolated and subcloned into pBluescript SK using Sali and

NotI restriction enzymes."

NotI restriction enzymes."

BASE COUNT 120 a 101 c 120 g 119 t

ORIGIN

Query Match 65.9%; Score 17.8; DB 43; Length 465;

Best Local Similarity 90.5%; Pred. No. 1.3e+02;

Matches 19; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 5 TTGTTCTACCACTGAAGTGG 25

I | | | | | | | | | | | | | | | | | | | | |

Db 151 TTGTTCCACCCTGAAGTGG 131

```

RESULT      8
T05736
LOCUS       T05736      337 bp      mRNA      EST      30-JUN-1993
DEFINITION  EST03625 Fetal brain, Strata gene (cat#936206) Homo sapiens cDNA
            clone HFBDF63 similar to Retrovirus-related pol polyprotein., mRNA
            sequence.
ACCESSION   T05736
NID         9316886
VERSION     T05736.1  GI:316886
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 337)
AUTHORS     Adams,M.D., Kerlavage,A.R., Fields,C. and Venter,J.C.
TITLE       3,400 expressed sequence tags identify diversity of transcripts
            from human brain
JOURNAL     Nature Genet. 4, 256-267 (1993)
MEDLINE     93364420
COMMENT     Other ESTs: EST03626
            Contact: Adams, MD
            The Institute for Genomic Research
            932 Clopper Road, Gaithersburg, MD 20878
            Tel: 3018699056
            Fax: 3018699423
            Email: mdadams@tigr.org
            Seq primer: M13-21.
FEATURES             location/Qualifiers
            source
            1..337
            /organism="Homo sapiens"
            /db_xref="ATCC (inhost):82427"
            /db_xref="taxon:9606"
            /clone="HFBDF63"
            /clone_lib="Fetal brain, Strata gene (cat#936206)"
            /note="Vector: LambdaZAP-II; 17-18 wk gestation, female;
            oligo-dr + random primed cDNA synthesis; lambdaZAP-II
            vector, 1.0kb average inser size."
BASE COUNT      100 a      70 c      68 g      94 t      5 others
ORIGIN
            1 ||||| ||||| ||||| ||||| |||||
Query Match      65.2%; Score 17.6; DB 20; Length 337;
Best Local Similarity 80.0%; Pred. No. 1.5e+02;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAGCTGTGTTCTACCACTGAACTAG 25
|| ||||| ||||| ||||| |||||
Db 137 AAACGTGTTCTAACACTTGTNCTAG 161

RESULT      9
F13213/c
LOCUS       F13213      308 bp      mRNA      EST      14-MAR-1995
DEFINITION  HSC3KCl11 normalized infant brain cDNA Homo sapiens cDNA clone
            c-3kcl1, mRNA sequence.
ACCESSION   F13213
NID         9709258
VERSION     F13213.1  GI:709258
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 308)
AUTHORS     Auffray,C., Behar,G., Bois,F., Bouchier,C., da Silva,C.,
            Devignes,M.D., Duprat,S., Houligatte,R., Juneau,M.N., Lamy,B.,
            Lorenzoni,F., Mitchell,H., Marlage-Samson,R., Pletu,G., Pouliot,X.,
            Sebastiani-Kabaktchis,C. and Tessier,A.
TITLE       IMAGE: molecular integration of the analysis of the human genome
            and its expression

```

JOURNAL MEDLINE COMMENT

C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
95277534

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1,rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

Single Read
Genexpress_library_idt: C; Genexpress_sequence_idt: ylc-3kcl1
Insert Length: 1240 Std Error: 0.00
Seq primer: (-21)M13_universal
High quality sequence stop: 298.
Location/Qualifiers

FEATURES Source

1..308
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-3kcl1"
/clone_lib="normalized infant brain cDNA"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate=muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 62 a 76 c 71 g 94 t 5 others

ORIGIN

Query Match 65.2%; Score 17.6; DB 21; Length 308;
Best Local Similarity 74.1%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 AAGCTGTGTTCTACCACTGAACTAGGC 27

Db 101 AAGCTGTTCNCCATGCANCAGGC 75

RESULT 10 T60600

LOCUS T60600 526 bp mRNA EST 13-FEB-1995
DEFINITION YB91905.r1 Strata gene liver (#937224) Homo sapiens cDNA clone
IMAGE:78584 5' similar to gb:K01763 HAPTOGLOBIN-1 PRECURSOR
(HUMAN);, mRNA sequence.

ACCESSION T60600

NID 9663637

VERSION T60600.1 GI:663637

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 526)

AUTHORS Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
Chiss-e,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W.,
Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N.,
Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L.,
Roifling,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J.,
Travaskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R.
and Marra,M.

Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478

TITLE JOURNAL MEDLINE COMMENT

Contact: Willson RK
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1027

High quality sequence stops: 283 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1027 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 283.

FEATURES

Location/Qualifiers

1..526
/organism="Homo sapiens"
/db_xref="GDB:498329"
/db_xref="taxon:9606"
/clone="IMAGE:78584"
/clone_lib="Stratagene liver (#937224)"
/sex="male"
/dev_stage="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"

/note="Organ: liver; Vector: pBluescript SK; Site_1: EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt. Hepatectomy from normal male caucasian. Average insert size: 1.1 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."

BASE COUNT 142 a 117 c 126 g 136 t 5 others
ORIGIN

Query Match 65.2%; Score 17.6; DB 21; Length 526;

Best Local Similarity 83.3%; Pred. No. 1.6e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAGCTTGTTCTACCACTGAACCTA 24

||||| ||||| ||||| ||||| |||||

Db 279 AAGCTTGTTCTACCACTGAACCTA 302

RESULT 11

AA004244/c

LOCUS

DEFINITION zh88406.r1 Soares fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA clone IMAGE:428363 5' similar to contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION AA004244

NID g1447919

VERSION AA004244.1 GI:1447919

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 425)

AUTHORS

Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags

JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478

COMMENT On May 18, 1995 this sequence version replaced gi:811621.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 673 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 395.

FEATURES

source

Location/Qualifiers
1..425
/organism="Homo sapiens"
/db_xref="GDB:1328132"
/db_xref="taxon:9606"
/clone="IMAGE:428363"
/clone_lib="Soares fetal_liver_spleen_INFLS_S1"
/sex="male"

/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: liver and spleen; Vector: p77T3D (Pharmacia) with a modified polylinker; Site_1: Pac I; Site_2: Eco RI; This is a subtracted version of the original Soares fetal liver spleen INFLS library. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTGAGAGATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p77T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 100 a 105 c 97 g 120 t 3 others
ORIGIN

Query Match 65.2%; Score 17.6; DB 27; Length 425;

Best Local Similarity 83.3%; Pred. No. 1.5e+02;

Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 AAGCTTGTTCTACCACTGAACCTA 24

||||| ||||| ||||| ||||| |||||

Db 83 AAGCTTGTTCTACCACTGATCAA 60

RESULT 12

AA004527/c

LOCUS

DEFINITION zh91a09.r1 Soares fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA clone IMAGE:428632 5' similar to contains Alu repetitive element;contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION AA004527

NID g1448104

VERSION AA004527.1 GI:1448104

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 480)

AUTHORS

Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Trevaskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.

TITLE Generation and analysis of 280,000 human expressed sequence tags

JOURNAL Genome Res. 6 (9), 807-828 (1996)

MEDLINE 97044478

COMMENT On May 9, 1995 this sequence version replaced gi:802992.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the


```

Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
    |||||
Db 1 TTACATATGAGCCTTCCATG 20

RESULT 2
HSII1AG/c
LOCUS Human gene for interleukin 1 alpha (IL-1 alpha). 24-APR-1993
DEFINITION X03833
ACCESSION X03833
NID G33785
VERSION X03833.1 GI:33785
KEYWORDS Alu repetitive sequence; interleukin 1 alpha; inverted repeat;
repetitive sequence.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 11970)
AUTHORS Fututani,Y., Notake,M., Fukui,T., Ohue,M., Nomura,H., Yamada,M. and
Nakamura,S.
TITLE Complete nucleotide sequence of the gene for human interleukin 1
alpha
JOURNAL Nucleic Acids Res. 14 (8), 3167-3179 (1986)
MEDLINE 86205226
REMARK Erratum:[[Published erratum appears in Nucleic Acids Res 1986 Jun
25;14(12):5124]]
COMMENT Data kindly reviewed (10-NOV-1986) by Y. Fututani.
FEATURES
    source
        location/Qualifiers
            1..11970
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            254..393
            /note="Alu repetitive sequence"
        misc_feature
            1375..1382
            /note="direct repeat 1"
        repeat_region
            1375..1390
            /note="pot. transcription activator (seq. homolog. to
            adenovirus 2 major late promoter transcription factor
            (MLTF) binding site)"
        repeat_unit
            1375..1382
            /note="inverted repeat A"
        repeat_unit
            1383..1390
            /note="inverted repeat A'"
        repeat_region
            1383..1390
            /note="direct repeat 1"
        TATA_signal
            1407..1413
        prim_transcript
            1438..11643
        exon
            1438..1488
            /number=1
        mRNA
            join(1438..1488,2153..2207,3166..3214,4103..4325,
            6262..6432,7815..7939,10290..11643)
        intron
            1489..2152
            /number=1
            2153..2207
            /number=2
        CDS
            join(2161..2207,3166..3214,4103..4325,6262..6432,
            7815..7939,10290..10490)
            /codon_start=1
            /product="IL-1-alpha"
            /protein_id="CAA27448.1"
            /db_xref="PID:G33786"
            /db_xref="GI:33786"
            /db_xref="SWISS-PROT:P01583"
            /translation="MAKVPDMPEDLKNYCSNEEDSSSIDHLSNQKSFYHVSYGPLH
            EGCMDQSVLSISTSKTFLKESWVVTATNGKVLKRLSLQSITDDLEAIAN
            DSEELIKPRSPFSLSKNYFNMRILIKYEFILNALNQSIIRANDQYLTAALHNL
            DEAYKFMGAYKSKDAKITVILRIKTYLTVAQDDQPVLLKEMPEIKTITGSE
            TNLFFWETHGTGKNYF"SVAHPNLFIATKQDYVWVCLAGGPPSITDFQILENQA"
            2208..3165
            /number=2
        intron

```

```

exon
    3166..3214
    /number=3
intron
    3215..4102
    /number=3
exon
    4103..4325
    /number=4
intron
    4326..6261
    /number=4
misc_feature
    4893..5174
    /note="Alu repetitive sequence"
exon
    6262..6432
    /number=5
intron
    6433..7814
    /number=5
misc_feature
    7695..7744
    /note="poly [dA-dC] tract"
exon
    7815..7939
    /number=6
intron
    7940..10289
    /number=6
misc_feature
    8466..8483
    /note="poly(dA-dC) tract"
repeat_region
    8912..9137
    /note="5 x 46 bp repeat"
misc_feature
    9770..9806
    /note="poly [dT-dG] tract"
exon
    10290..11643
    /number=7
polyA_site
    11643
    /note="polyA site"
misc_feature
    11863..11970
    /note="Alu repetitive sequence"
BASE COUNT 3708 a 2489 c 2226 g 3547 t
ORIGIN

Query Match 92.0%; Score 18.4; DB 10; Length 11970;
Best Local Similarity 95.0%; Pred. No. 2.6;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
    |||||
Db 569 TTACATATGAGCCTTCAATG 550

RESULT 3
HSDJ925J7/c
LOCUS Homo sapiens chromosome 22 clone DJ925J7, WORKING DRAFT SEQUENCE,
DEFINITION in unordered pieces.
ACCESSION AL078622
NID G5051357
VERSION AL078622.1 GI:5051357
KEYWORDS HTG; HTGS-PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 118605)
AUTHORS Matthews,L.
TITLE Direct Submission
JOURNAL Submitted (11-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquerry@sanger.ac.uk
COMMENT This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dJ925J7 Contig_ID: 01919 acc-AL078622
Length: 13921 bp Unfinished: dJ925J7 Contig_ID: 02005
acc-AL078622 Length: 16691 bp Unfinished: dJ925J7 Contig_ID:

```

Matches	17;	Conservative	0;	Mismatches	1;	Indels	0;	Gaps	0;
---------	-----	--------------	----	------------	----	--------	----	------	----

COMMENT

requests: clonerequest@sanger.ac.uk
 On Jan 16, 1999 this sequence version replaced g1:3927943.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence is the entire insert of clone 596H12. This sequence
 has been finished according to sequence map criteria as follows. An
 attempt is made to resolve all sequencing problems, such as
 compressions and repeats, but not necessarily within known
 annotated human repeat sequence elements (e.g. Alu). Where the
 sequence is ambiguous, there is an annotation using the 'unsure'
 feature key.
 This sequence was generated from part of bacterial clone contigs of
 human chromosome 6, constructed by the Sanger Centre Chromosome 6
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr6>
 596H12 is from the library RPC14 constructed at the Roswell Park
 Cancer Institute by the group of Pieter de Jong. For further
 details see <http://bacpac.med.buffalo.edu/VECTOR:pcYPAC2>.

FEATURES

source	1. .175555 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /clone="596H12" /map="p22.2-22.3" /clone_lib="RPC14" 710. .840 /note="Trigger3(Golem) repeat: matches 2. .142 of consensus" 838. .1095 /note="Trigger3(Golem) repeat: matches 2781. .3038 of consensus" 1540. .1928 /note="Ltr16A repeat: matches 18. .440 of consensus" 2230. .2825 /note="HERV16 repeat: matches 328. .940 of consensus" 2910. .3374 /note="Ltr16A repeat: matches 5669. .6146 of consensus" 3787. .4161 /note="Ltr16A repeat: matches 5771. .6142 of consensus" 4167. .4545 /note="HERV16 repeat: matches 1571. .1944 of consensus" 6293. .6503 /note="HERV16 repeat: matches 4443. .4661 of consensus" 7087. .7374 /note="AluJb repeat: matches 1. .289 of consensus" 7492. .7746 /note="Ltr16A repeat: matches 25. .291 of consensus" 7913. .8039 /note="Ltr16A repeat: matches 317. .442 of consensus" 8762. .8831 /note="MSTA repeat: matches 361. .426 of consensus" 8832. .9142 /note="AluSg repeat: matches 1. .313 of consensus" 9143. .9475 /note="MSTA repeat: matches 1. .361 of consensus" 9222. .9706 /note="match: GSS AQ219540" 9884. .10190 /note="AluSg repeat: matches 1. .308 of consensus" complement(11134. .11527) /note="match: GSS AQ086021" complement(11525. .11848) /note="match: GSS AQ062320" 12210. .12442 /note="MIR repeat: matches 8. .262 of consensus" 12780. .13027 /note="LIME1 repeat: matches 5533. .5788 of consensus" 14101. .14503 /note="L2 repeat: matches 972. .1381 of consensus"
repeat_region	14703. .15322 /note="L2 repeat: matches 2041. .2745 of consensus" 15692. .15922 /note="MSTB repeat: matches 234. .426 of consensus" 15923. .16196 /note="AluSg repeat: matches 1. .276 of consensus" 16197. .16416 /note="MSTB repeat: matches 2. .234 of consensus" complement(17504. .175317) /note="match: EST R02441" 17689. .17822 /note="MIR repeat: matches 123. .256 of consensus" 19121. .19677 /note="LIME1 repeat: matches 5127. .5673 of consensus" complement(19601. .19957) /note="match: STS 224329: genomic marker D6S461: polymorphic TC, CA and TA repeats in 19687. .19853" 19686. .19835 /note="75 copies 2 mer gt 82% conserved" 20167. .20548 /note="LIME1 repeat: matches 5774. .6151 of consensus" 20989. .21102 /note="MER81 repeat: matches 1. .114 of consensus" 21255. .21347 /note="MIR repeat: matches 168. .261 of consensus" 22169. .22944 /note="L2 repeat: matches 1898. .2746 of consensus" 23048. .23230 /note="MIR repeat: matches 29. .214 of consensus" 23563. .23740 /note="MER91A repeat: matches 1. .183 of consensus" 24960. .25161 /note="MER2 repeat: matches 8. .210 of consensus" complement(25042. .25379) /note="match: GSS AQ191276" 25473. .25759 /note="AluSx repeat: matches 1. .287 of consensus" 26467. .26540 /note="L2 repeat: matches 2237. .2321 of consensus" 26573. .26786 /note="L2 repeat: matches 2314. .2522 of consensus" 27198. .27348 /note="MTR1J repeat: matches 332. .496 of consensus" 28127. .28410 /note="AluSg repeat: matches 1. .286 of consensus" 28995. .29065 /note="MSTB repeat: matches 355. .426 of consensus" 29066. .29361 /note="AluJ repeat: matches 1. .296 of consensus" 29362. .29723 /note="MSTB repeat: matches 1. .355 of consensus" 31269. .31391 /note="L2 repeat: matches 2638. .2750 of consensus" 33043. .33380 /note="MER57B repeat: matches 64. .400 of consensus" 33076. .33435 /note="MER93 repeat: matches 1. .361 of consensus" 33945. .34322 /note="match: GSS B87850" 35631. .35737 /note="L2 repeat: matches 2632. .2738 of consensus" complement(36270. .36631) /note="match: GSS AQ182323" 37479. .37644 /note="MTR1J repeat: matches 22. .181 of consensus" 37689. .37776 /note="MER5A repeat: matches 4. .87 of consensus" 37898. .38022 /note="MER5A repeat: matches 55. .189 of consensus" 38042. .38099 /note="MER5B repeat: matches 120. .173 of consensus" 38100. .38827 /note="Ltr8 repeat: matches 1. .691 of consensus"

```

repeat_region 38828..38873
/Note="MER583 repeat: matches 62..120 of consensus"
repeat_region 40174..40709
/Note="MER41C repeat: matches 9..554 of consensus"
repeat_region 40720..40879
/Note="LTR16B repeat: matches 137..304 of consensus"
repeat_region 40927..41658
/Note="MER50 repeat: matches 1..722 of consensus"
repeat_region 42300..42409
/Note="MER58B repeat: matches 69..178 of consensus"
repeat_region 42536..43325
/Note="L1MC1 repeat: matches 5555..6325 of consensus"
repeat_region 43982..44194
/Note="MIR repeat: matches 28..262 of consensus"
repeat_region 45047..45203
/Note="FRAM repeat: matches 1..167 of consensus"
repeat_region 45278..45477
/Note="L1MC4 repeat: matches 7308..7511 of consensus"
repeat_region 45478..45843
/Note="THE1C repeat: matches 1..371 of consensus"
repeat_region 45844..46192
/Note="THE1C-internal repeat: matches 1..350 of consensus"
repeat_region 46198..47354
/Note="THE1C-internal repeat: matches 427..1580 of consensus"
repeat_region 47355..47733
/Note="THE1C repeat: matches 1..371 of consensus"
repeat_region 47817..48057
/Note="L1MD3 repeat: matches 7484..7739 of consensus"
repeat_region 49840..49883
/Note="22 copies 2 mer gt 93% conserved"
repeat_region 51669..51763
/Note="L2 repeat: matches 2628..2708 of consensus"
repeat_region 52271..52350
/Note="L1MC repeat: matches 1593..1674 of consensus"
repeat_region 52472..52566

```

QY 2 TACATATGAGCCTTCCAT 19

Db 54844 TACATATGAGCATTCCAT 54827

```

RESULT 8
AC005884/c
LOCUS
DEFINITION Homo sapiens chromosome 17, clone hRPK.264_B_14, complete
sequence.
ACCESSION AC005884
NID 3858902
VERSION AC005884.1 GI:3858902
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 120187)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
TITLE Homo sapiens chromosome 17, clone hRPK.264_B_14
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 120187)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,

```

Mychaleckyj,J., Nafz,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Tichovolisky,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolisky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.

TITLE

JOURNAL

Submitted (28-OCT-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

REFERENCE

AUTHORS

3 (bases 1 to 120187)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C.,
Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Gardyna,S., Geraigery,K., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L.,
Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P.,
McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Nafz,R., Naylor,J., Niloff,M., O'Connor,T.,
O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stilwell,J., Tichovolisky,N., Stone,C.,
Subramanian,A., Tesfaye,S., Tichovolisky,N., Torruella-Miller,I.,
Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,
Ye,W.J., Zhao,J. and Zody,M.

TITLE

JOURNAL

Submitted (10-NOV-1998) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

COMMENT

On Nov 10, 1998 this sequence version replaced gi:3850604.
All repeats were identified using RepeatMasker: Smit, A.F.A. &
Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Only the last 120187 base pairs of this clone are being submitted.
The remainder overlaps with accession number AC005856(WICGR project
L452).

FEATURES

source

Location/Qualifiers

```

1..120187
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone=" hRPK.264_B_14"
/map="17"
/chromosome="17"
136..253
/rpt_family="MIR"
1699..1783
/rpt_family="MIR"
complement(1997..2070)
/rpt_family="L2"
complement(3529..3821)
/rpt_family="AluY"
4933..5011
/rpt_family="MIR"
5197..5497
/rpt_family="AluY"
complement(5510..5833)
/rpt_family="L1M4c"
5844..6096
/rpt_family="L1ME3"
complement(6097..6311)
/rpt_family="AluJb"
6312..6497
/rpt_family="L1ME3"
6524..6663
/rpt_family="AluSg/x"
6671..6776
/rpt_family="L1ME3"
complement(6792..6820)
/rpt_family="(CA)n"
complement(6888..6898)
/rpt_family="Charliel"
6899..7192

```

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

repeat_region

```
repeat_region /rpt_family="AluX"
repeat_region complement(7193..7810)
repeat_region /rpt_family="Charlie1"
repeat_region complement(7811..8099)
repeat_region /rpt_family="AluSp"
repeat_region complement(8100..8738)
repeat_region /rpt_family="Charlie1"
repeat_region 8762..9063
repeat_region /rpt_family="L1ME3"
repeat_region complement(9064..9357)
repeat_region /rpt_family="AluJb"
repeat_region 9358..9457
repeat_region /rpt_family="L1ME3"
repeat_region 9458..9765
repeat_region /rpt_family="AluX"
repeat_region 9766..10787
repeat_region /rpt_family="L1ME3"
repeat_region 11711..11990
repeat_region /rpt_family="AluSg"
repeat_region 11991..12016
repeat_region /rpt_family="(GAAA)n"
repeat_region 12834..13147
repeat_region /rpt_family="AluX"
repeat_region complement(14452..14480)
repeat_region /rpt_family="(CAAAA)n"
repeat_region complement(14481..14764)
repeat_region /rpt_family="AluX"
repeat_region 16273..16576
repeat_region /rpt_family="AluX"
repeat_region 18854..19070
repeat_region /rpt_family="AluSc"
repeat_region complement(19232..19308)
repeat_region /rpt_family="L2"
repeat_region 20138..20305
repeat_region /rpt_family="L1MC5"
repeat_region 20314..20446
repeat_region /rpt_family="L1MB6"
repeat_region complement(21891..22202)
repeat_region /rpt_family="AluJb"
repeat_region complement(22235..22460)
repeat_region /rpt_family="L2"
repeat_region 22461..22782
repeat_region /rpt_family="AluJo"
repeat_region 22785..23053
repeat_region /rpt_family="AluX"
repeat_region 23054..23085
repeat_region /rpt_family="(TAAAA)n"
repeat_region complement(23086..23144)
repeat_region /rpt_family="L2"
repeat_region 24563..24608
repeat_region /rpt_family="(CA)n"
repeat_region 24876..24901
repeat_region /rpt_family="AT_rich"
repeat_region complement(25113..25147)
repeat_region /rpt_family="(TAAAA)n"
repeat_region complement(25148..25429)
repeat_region /rpt_family="AluY"
repeat_region 25493..25536
repeat_region /rpt_family="AT_rich"
repeat_region 26609..26787
repeat_region /rpt_family="AluJo"
repeat_region 27051..27337
repeat_region /rpt_family="AluSg"
repeat_region 27426..27478
repeat_region /rpt_family="(TA)n"
repeat_region complement(28432..28519)
repeat_region /rpt_family="L1MC1"
repeat_region complement(28533..28837)
repeat_region /rpt_family="AluJo"
repeat_region complement(29747..29960)
repeat_region /rpt_family="L2"
repeat_region 30103..30177
repeat_region /rpt_family="L2"
```

```
repeat_region 30257..30440
repeat_region /rpt_family="L1MC5"
repeat_region 31072..31236
repeat_region /rpt_family="MERSA"
repeat_region complement(31697..31726)
repeat_region /rpt_family="(CAA)n"
repeat_region 32265..32298
repeat_region /rpt_family="(TA)n"
repeat_region complement(32810..33109)
repeat_region /rpt_family="AluX"
repeat_region 33695..33716
repeat_region /rpt_family="AT_rich"
repeat_region 34886..34923
repeat_region /rpt_family="(TA)n"
repeat_region complement(34924..35228)
repeat_region /rpt_family="AluSc"
repeat_region 35229..35260
repeat_region /rpt_family="(TA)n"
repeat_region 35344..35409
repeat_region /rpt_family="(TA)n"
repeat_region complement(38342..38515)
repeat_region /rpt_family="FAM"
repeat_region 38608..38635
repeat_region /rpt_family="HSMAR2"
repeat_region 38632..39098
repeat_region /rpt_family="HSMAR2"
repeat_region 39099..39409
repeat_region /rpt_family="AluY"
repeat_region 39410..40054
```

Query Match 82.0% Score 16.4; DB 11; Length 120187;
Best Local Similarity 94.4% Pred. No. 41;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCA 18
|||||
Db 74248 TTACTTATGAGCCTTCCA 74231

RESULT 9
CMU17165/c

LOCUS CMU17165 973 bp mRNA ROD 04-DEC-1995
DEFINITION Cricetulus migratorius Ig kappa light chain mRNA, complete cds.
ACCESSION U17165
NID 9841147
VERSION U17165.1 GI:841147
KEYWORDS Armenian hamster.
SOURCE Cricetulus migratorius
ORGANISM Cricetulus migratorius
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae; Cricetulus.
REFERENCE
AUTHORS Whitters,M.J. and Collins,M.
TITLE Hamster cDNA homologs to the mouse immunoglobulin kappa constant
and Igk-v 45.1 genes
JOURNAL Immunogenetics 42 (3), 227-228 (1995)
MEDLINE 95369852
REFERENCE 2 (bases 1 to 973)
AUTHORS Collins,M.
TITLE Direct Submission
JOURNAL Submitted (15-NOV-1994) Mary Collins, Genetics Institute, 87
Cambridge Park Dr., Cambridge, MA 02140, USA
FEATURES
source
1..973
/organism="Cricetulus migratorius"
/db_xref="taxon:10032"
/dev_stage="rearranged"
50..769
/codon_start=1
/product="immunoglobulin kappa light chain"
/protein_id="AA82732.1"
/db_xref="PID:9841148"
/db_xref="GI:841148"

/translation="MKLPVLLALLLFIPIPSRGDVMVTQSPNVLSVSLGEOVSISCR
SSQSLVQNGNTYVNMFLQRPQSPKRLIYIKTSNRNSGVPDFTFSGSGDKDFLTKISR
METEDGYVYCMQGSYVPWTFPGTKLEIKRADAKPTVSIFFPSGEOLGTGSAITLVCF
VNNFYPKIDINVKWKYDGEKRDGVLSQVTDQDSKDYSLSLSTLSLTRADYERHNYLT
CEVTHKSTAAIVKTLNRNEC"
V_region 50..391
J_segment 392..448
C_region 449..766
BASE COUNT 267 a 246 c 219 g 241 t
ORIGIN

Query Match 82.0%; Score 16.4; DB 12; Length 973;
Best Local Similarity 94.4%; Pred. No. 32;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 ACATATGAGCCTTCCATG 20
|||||
Db 405 ACATATGAGCCTTGCATG 388
|||||

RESULT 10
AF098256 450 bp DNA BCT 15-NOV-1998
LOCUS
DEFINITION Pseudomonas syringae pv. pisi strain 539 internal transcribed
spacer 1, complete sequence.
ACCESSION AF098256
NID 93873271
VERSION AF098256.1 GI:3873271
KEYWORDS
SOURCE Pseudomonas syringae pv. pisi.
ORGANISM Pseudomonas syringae pv. pisi
Bacteria; Proteobacteria; gamma subdivision; Pseudomonas group;
Pseudomonas.

REFERENCE 1 (bases 1 to 450)
AUTHORS Koike,S.T., Barak,J.D., Henderson,D.M. and Gilbertson,R.L.
TITLE Bacterial blight of leek: A new disease in California caused by
Pseudomonas syringae
JOURNAL Plant Dis. (1998) In press
2 (bases 1 to 450)
AUTHORS Barak,J.D., Gilbertson,R.L. and Koike,S.T.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Plant Pathology, University of California,
Davis, One Shields Blvd., Davis, CA 95616, USA
Location/Qualifiers

FEATURES
Source
1..450
/organism="Pseudomonas syringae pv. pisi"
/strain="539"
/db_xref="taxon:59510"
/note="pathovar: pisi"
misc_RNA
1..450
/note="ITS1"
/product="internal transcribed spacer 1"
BASE COUNT 106 a 110 c 113 g 121 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 2; Length 450;
Best Local Similarity 89.5%; Pred. No. 69;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCAT 19
|||||
Db 324 TTACAAATGAGCATTCAT 342
|||||

RESULT 11
SC41KCIV 41175 bp DNA PLN 10-NOV-1995
LOCUS
DEFINITION S.cerevisiae DNA (Cosmid 31A2; chromosome IV; 41 kb).
ACCESSION 267750
NID g1061256
VERSION 267750.1 GI:1061256

KEYWORDS
SOURCE ORGANISM

REFERENCE 1 (bases 1 to 41175)
AUTHORS Pohl,T.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 41175)
AUTHORS Pohl,T.M.
TITLE Direct Submission
JOURNAL Submitted (09-NOV-1995) Thomas M. Pohl, GATC GmbH,
Fritz-Arnold-Str. 23, Konstanz, 78467, Germany
Location/Qualifiers

FEATURES
Source
1..41175
/organism="Saccharomyces cerevisiae"
/strain="alphas288C"
/db_xref="taxon:4932"
/chromosome="IV"
/clone="cosmid 31A2:"
/clone_lib="cosmid library (C. Jacq, ENS, France)"
CDS
1..766
/codon_start=2
/product="putative protein"
/protein_id="CAA91564.1"
/db_xref="PID:g1061257"
/db_xref="GI:1061257"
/db_xref="SWISS-PROT:P48570"
/translation="IGVNRVGIADTVGCANPQVYELIRTLKSVVSCDIECHFNHDTG
CALANAYTALEGGARLIDVSLVIGIGERNGIPLGGLMAMIVAAADYVSKYKLLKHIR
DIENLVADAEVNIPIFNPIITGFCFATHKAG:HKAKILANPSTYIILPHDPGMKRYI
HFANRLTGMNAIKARVDQLNLNLTDDQIKVETAKIKLGDSRLNIDDDVDSDIKNFHA
EYSTPQVLSAKKNKNSDVPDLATIPAAKRTKPSA"
1939..3582
/codon_start=1
/product="putative protein"
/protein_id="CAA91565.1"
/db_xref="PID:g1061258"
/db_xref="GI:1061258"
/db_xref="SPTREMBL:Q12301"
/translation="MVRLNHAASYFMPICFSTRPHIVILSALFSISLSFVYASSELL
LHQVDDPMFNPNSQDIYETELGLFSPIYFLKTLFNLNQRELIINLIVDFPIND
VFMLLILGLAYPOVDHEGGT:IKHKECSWHIIPROYIFGISWALGETTICIGNLF
NYQEIADPNINSGETFHESANTYCNNDMSHNDCCGCTEYRPNVVDSDITLSKIE
VRNDSISNNVYSEYHPKIPLRSSSTYGSIROQPHENKCOLHVPNDSQDITIMM
NPIDNSKLTLTLDGDSFPIDEOPILKKSFGYTWAVNENTONTTKSFTPIKRFIA
FSTAYOLVTGLLLMILVGSNIMLTIGESLLSILSMFYVVRGHEGLFTPVVTFGSKTI
SNFLCVLIPISLNFNLNTSIYLRREDDMFNNSQGFEDDENTISKRVAITNOEYQ
HPLSANYISMDSQPDVINSPPGHFGMNSQGLLGNLTLYYGSGLNGDDDDMTNDSALLRFC
KKLVKNRALARNSDFVLGVVMVSNLSLVFTGILSTVYI"
4334..5248
/codon_start=1
/product="putative protein"
/protein_id="CAA91566.1"
/db_xref="PID:g1061259"
/db_xref="GI:1061259"
/db_xref="SPTREMBL:Q12477"
/translation="MISDYDALOFNKKPVSOEMIOFLATSTASLIKIRENNNPQGC
RPDLSIFIKNVVIOQSNVQPTLMAATSVYLNKLSVIEKPNVYGINTRHRIFLGCLIL
AAKTLNDSPPNKKHTWTTEGLRLREVTNTERLELLEYNMDVRITTPDLIDLSLYFL
GP:IKEQFLQRRQEMLEFNAPSQOLKEYINRRPVSHSRTSSAISVPSLTSMAVST
TDSRSSLLAKYQPSLPVESDNFNKKHNVLRNNDICNFRAEENIHVSNNHIDVTMG
SSPVMNSHKPTIQRNLNFTRRGSSFFKQ"
5565..7157
/codon_start=1
/product="putative protein"
/protein_id="CAA91567.1"
/db_xref="PID:g1061260"
/db_xref="GI:1061260"
/db_xref="SWISS-PROT:P46681"
/translation="MLRNILVRSIGSNFKFAGRYNKKSSALLGYRVRVNYSTKTIQTRL

CDS
/codon_start=1
/product="putative protein"
/protein_id="CAA91566.1"
/db_xref="PID:g1061259"
/db_xref="GI:1061259"
/db_xref="SPTREMBL:Q12477"
/translation="MISDYDALOFNKKPVSOEMIOFLATSTASLIKIRENNNPQGC
RPDLSIFIKNVVIOQSNVQPTLMAATSVYLNKLSVIEKPNVYGINTRHRIFLGCLIL
AAKTLNDSPPNKKHTWTTEGLRLREVTNTERLELLEYNMDVRITTPDLIDLSLYFL
GP:IKEQFLQRRQEMLEFNAPSQOLKEYINRRPVSHSRTSSAISVPSLTSMAVST
TDSRSSLLAKYQPSLPVESDNFNKKHNVLRNNDICNFRAEENIHVSNNHIDVTMG
SSPVMNSHKPTIQRNLNFTRRGSSFFKQ"
5565..7157
/codon_start=1
/product="putative protein"
/protein_id="CAA91567.1"
/db_xref="PID:g1061260"
/db_xref="GI:1061260"
/db_xref="SWISS-PROT:P46681"
/translation="MLRNILVRSIGSNFKFAGRYNKKSSALLGYRVRVNYSTKTIQTRL

TSNYPDVHRRPREFKLLSDLNFKYSIISEQEILRASESEDLSFYNEDMMRKYGQS
KLVLRPKSVKSLIILNYCNDEKIAVYPOGGNTGLVGGVPIFDELILSLANLKIRD
FDVSGILKCDAGVILANNVMEQNIYMFPLDLAGKSGHVGGVATNAGGURLLRY
GSLHGVGLGVPMNGOIVNNSMRKONTGYDLKQLFIGSEGTGIITGVSLTVP
KPAFNVSILSVESFEDVQVYFRAQELSEILSAFEMDAKSQVLAKSOLKDAEPL
EDHPFILLTETSGNKHDDHDSKLETFLENNMEGIVTDGVVAQDETQNLQNGVHGD
IPASQAANGYKIDYVSLPKUDILSYLEATNARLSBAELVDPSPKVPVGAIGYGHWD
GNLHNAVREYKNIENKTLEPPYFVFSVSKHGSVAEHLGLGQKKNYIGYSKSPEV
KMKDKLVHYPNGILNPYKI*
complement(7252..7764)

CDS

/codon_start=1
/product="putative protein"
/protein_id="CAA91568.1"
/db_xref="PID:g1061261"
/db_xref="GI:1061261"
/db_xref="SWISS-PROT:Q12257"
/translation="MSKNVGLKLVKINNESEVLVDRKSKTOARCCPLQONOKDIPSILQE
LTQNNKSVKASHWHYAWRTAEVSNLHQQKKGNKANKSNHSHNSNIYVQ
PKNIEQACGGEAGAGQRLTLTLLERANIFNLVI VTRWYGGTPLGSSRFRHISTCAV
ETLKKGGFLP*
8140..10266

CDS

/codon_start=1
/product="putative protein"
/protein_id="CAA91569.1"
/db_xref="PID:g1061262"
/db_xref="GI:1061262"
/db_xref="SPTREMBL:Q12027"
/translation="MATGRIQFAVSTPCNTKGPSPGYRLEFEKNDRLALVPSRGCTK
VDVNAIQAFCYLRPNCRDTSISPDAITHLSDCDYMLAKSNGFIETISNYQYKIKNG
LRLAPSILCTPEDFESNFDYMTAGLEYSOGLLYCCMSCRIYFVNNLPTDIYQ
YKMNIPMFDCFKVHHNDNTHSSEKLFEGSTRYIGRSGSKHCYICLLPIEPSH
LRSPVSVSCNNYQGLPIYRSPYMLHIERGISTFHINPLDRFCFTVSPRSLFIRK
IILPVTYLFSTFISLKSIOGDTCEILSWNVAQONGFSLFSWISNKFETDID
INSTIMWDIVKYSQGTMLDQAFVIMKQGHAKDDIYELFHTQDMLGSSNNSFSFAS
SEPRLSRRRESFOALTDAFERMDVPCSTKWLDSFIRGLRNTTSMYDVEIIEKI
SHNGNDVNEEDNTDSEDTWSETLDNKKMDIWCIDHFVTLNAPFRKYDEPII
KIDSLNKNSENGTNEEWAESQMKVGGVVIDETAQKALNGNCSFKFLWLODS
LCFLDTHGVLNINRFPIKTKNLLRNSKOTIRIIPHDGFLINDTIVINDIDGTDN
VCAALRFLDYDDEKLTFFIVKVKRD*
complement(10527..11561)

CDS

/codon_start=1
/product="putative protein"
/protein_id="CAA91570.1"
/db_xref="PID:g1061263"
/db_xref="GI:1061263"
/db_xref="SPTREMBL:Q12476"
/translation="MEKNTAPFVVDVDTPTPDKLVAPSTEENVSNPNELRALRGGR
YFGVSDDDKDAIKKAPKCNCSQORGLKKDCPHLICSYCGATDDHYSRCPKAIQCS
KDEVHYRQCQPHKMKVOCITLCKSKHKRCPSIWRAYIILVDDNEKAKPKVLPFH
TIYVCGGKGHFGDDCKEKRSRVPNEDGSFTGSLNVELKOEYRHMNRNSDNE
DYQFSEIYDEPLPRPSHRHSONSHSRNKRASNPFPYQKSNVIOPTIRGE
TSLNNNISKNSRYQNTKVNVSSTSENMYGSRYNPSTYVDNNISNSSNRYNRYSTP
YRSGTLCKRR*
complement(11869..13632)

CDS

/codon_start=1
/product="D-lactate dehydrogenase"
/protein_id="CAA91571.1"
/db_xref="PID:g1061264"
/db_xref="GI:1061264"
/db_xref="SWISS-PROT:P32891"
/translation="MLWKRTCTRLIKPIAQRGLVRRRCYRYASTGTGSTDSSQWL
KYSVIASSATLFGVLFKAKNLYSRETKEDLIEKLEVMKVIDPVNSTLKLSSLDSPYLH
DPVKIDRWELDKVQGNKPNYSDAKSLDASHDYFNTFHPSPQORPIILFPHIT
EESVKILCHDNNMPVPSGGTSLGELHPLTRIGDITVDLSKFMNNVVKFKDLID
DITVOAGLPWHDNYLSDHGLMFGDPPGAOIGGCIANSCGTNAYRYGTWKENII
NMITLVDGTIVTKKPRKSSAGYNLGLFVSGEFTLGIVTEATVKCHVKPKAEVVA
WYSEDTIKDAACASNLPTQSGIHLNMLLEDNMMKLINASESTDRCDDWKEKTFMFK
IGRSPVINALDEVKVAQLNHCNSFQAKDDDEKLELWEAKKVALWVLDADSK
DKSAKITVDVAVPSQDFVRIHETKDMQASKLINAIVGHGAGDNFHFPIVYRTPEE
HETCSOLVDRMKRALNAEGTCTGEGHGVIGKREYLLLEELGEAPVDMLRKIKLAIDPK
RIMNPDKIFKTDNPNANDYR*
14234..15121

CDS

/codon_start=1
/product="putative protein"
/protein_id="CAA91572.1"
/db_xref="PID:g1061265"
/db_xref="GI:1061265"
/db_xref="SPTREMBL:Q12515"
/translation="MATFNPQMEMENQARVOEYKVSSTGRGGAGNIHKSMSKPSVLLP
LKSNGTAVANNNGNSQEKVPRFAIGRGAGNIHFDPHLTRSAQQLDSNDNINNDYD
INDIDDYISPIITSDMYDEGSPNPTNRSRISATRSRHSQSLHATTSPPNNAPATVVRG
GAGNIFNKKKVASNGNEEDTIRGNIIEDEDTINANEDNLFVVTSGNALAAIKSTS
KPKKKLKGSAPKFAIGRGAGNIISPKSSRNTNHNLNDDDDDEKFNLKDDNGREK
KKKKKKSGFFSSLKTMFN*
complement(14650..15129)
/codon_start=1
/product="putative protein"
/protein_id="CAA91573.1"
/db_xref="PID:g1061266"
/db_xref="GI:1061266"
/db_xref="SPTREMBL:Q12135"
/translation="MLFQNLIVFKELKNPDEFFFFFSSFLPSLSSFKNLSSSSSFKLW
LIVFLDFGEIMFPAPPLPIANFSGALLPLSLFLGFLDVLIAAKALPLEVTTNRLSS
FALIVSSSSILPPRISSSPFPLLATFFLLKKMFAPPPIPTTIGALLFGDEVVACRD*
complement(15246..21683)

CDS

/codon_start=1
/product="putative protein"
/protein_id="CAA91574.1"
/db_xref="PID:g1061267"
/db_xref="GI:1061267"
/db_xref="SWISS-PROT:Q12680"
/translation="MPVLKSDNFDPLEEAYEGGTIONYNDEHLHLSWANVIPDKRGL
XPDYEDHACGVGVANKGSEQSHKIVTDARYLLVNNTHRGAVSSDNGDGGAGILIGI
PHEFKREFKLDLIDPEMGYAVGNVFFKNEKNKNLICKQKIFEDLAASFNL
VLGWRNPVDSITLGDVALSREPTILOPLLPLYDEKQEFNETKERTOLYLLRKEAS
LOIGLENWFVCSLNTITVYKQLTPAOVYVYVYDITNAHFKSHALVHSRSTNTF
PSWDRAPRLWLAHGEINTLRGNKNMRSREGVMNSATFKDELKALYPIIEGGSDS
AALDNVELLTINGLUSLPAEVMMPYEAHKMDSDLRKADLAACLMPEWDPGALL
TDTDGYCGAILDRNGLRCPYITSDRIVICASEGVPIENSLVQKGLKPGDLF
LVDTQYGEVDTKKLSQISKRQDFKWSLKVILDDLLSKTANLVYKKEFISQDSL
KQVSDRLLANGYTFEQVTFILTPMALTKGKALGSMNDAPLACLNENPVLLDYFRQ
LFAQVTPPIDPIREANVMSLECYVCGQGNLLEMHSSQCDRLLLKSPILHWNFEOLK
NIEAAYPSVSAEIDITFDKSEGLLYTDTIDKTIKLAESAIDGKKLITDRKGA
NRVSISSIAISCIHHHLNRKORSVALILETGEAREIHHFCVLVLCGCDGVPIYLA
METLVNRMREGLLRVNNNDNTLEEGQILENYKHAIDAGILKVMNKMISTLAAYKA
QFIEALGDNISIVDLCTSSIRIGVTFEYLAQDAFSLHERGYPRQCTISKVNLPE
SGEYHFRDGGYKHVNEPTAISLQDTRNRKNDVSWOLYVKKMEARQCTLRGLLED
FENSVIPLOEVPWTEIARRFASGAMSYGSIISMEASHLIAIAMNRLGAKSNGEGE
DAERSAVENGDMTMSAIKQVASAREGVTSYLSLDADELCIKIAQCAKPGEGELPAH
KYSKDTAKTRHSTPNVGLISPPPHDIYSTEDLQILYDLKCANPRAGISVULSEVG
VGIVASGVAKAKADHILVSGHDGTTGAARWTSVKYAGLPWELGLAETHOTLVLDNRR

Query Match 79.0%; Score 15.8; DB 7; Length 41175;
Best Local Similarity 89.5%; Pred. No. 87;
Matches 17; C. servative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACATATGACCTTCATG 20
||||||| | |||||
Db 30398 TACATATGACCATTCATG 30416

RESULT 12
SCDC36
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
SCDC36 1208 bp DNA PLN 12-SEP-1993
Yeast cell division control gene CDC36.
X04287
g3490
X04287.1 GI:3490
CDC36 gene; cell division control.
baker's yeast.
Saccharomyces cerevisiae
Saccharomycetes
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomycetes.
1 (bases 1 to 1208)
Ferguson,J., Ho,J.Y., Peterson,T.A. and Reed,S.I.

TITLE	Nucleotide sequence of the yeast cell division cycle start genes CDC28, CDC36, CDC37, and CDC39, and a structural analysis of the predicted products			
JOURNAL	Nucleic Acids Res. 14 (16), 6681-6697 (1986)			
MEDLINE	86312926			
FEATURES	Location/Qualifiers			
source	1..1208			
gene	/organism="Saccharomyces cerevisiae" /db_xref="taxon:4932" 1..1208			
misc_feature	/gene="CDC36" /db_xref="SGD:S0002324" 24..29			
misc_feature	/gene="CDC36" /note="TATA-like sequence" /db_xref="SGD:S0002324" 77..81			
misc_feature	/gene="CDC36" /note="TATA-like sequence" /db_xref="SGD:S0002324" 145..147			
misc_feature	/gene="CDC36" /note="pot. start codon" /db_xref="SGD:S0002324" 157..160			
misc_feature	/gene="CDC36" /note="CAAT-like sequence" /db_xref="SGD:S0002324" 189..192			
misc_feature	/gene="CDC36" /note="TATA-like sequence" /db_xref="SGD:S0002324" 207..782			
CDS	/gene="CDC36" /note="CDC36 gene product (aa 1-191)" /codon_start=1 /db_xref="SGD:S0002324" /protein_id="CAA27835.1" /db_xref="PID:g3491" /db_xref="GI:3491" /translation="MKFGIKALVPLLLKLEDKELSSYDHSMTLGLADLSMLYSLGIP RDSQHRVLTDFOSPAETSRSEVPRFFPEFTNPGLVQSTVTPPCFNSIONDQO RVAFODETLFFLYFKHGPVFIQELVLELRKNWRYHKTLEKAWLTGKPMPEFIVASD GLSEGSYVFDFQREKCORDFLFYNAIM" 806..814			
misc_feature	/gene="CDC36" /note="pot. polyA signal" /db_xref="SGD:S0002324" 822..826			
misc_feature	/gene="CDC36" /note="pot. termination sequence 1" /db_xref="SGD:S0002324" 836..839			
misc_feature	/gene="CDC36" /note="pot. termination sequence 1" /db_xref="SGD:S0002324" 893..896			
misc_feature	/gene="CDC36" /note="pot. termination sequence 2" /db_xref="SGD:S0002324" 904..906			
misc_feature	/gene="CDC36" /note="pot. termination sequence 2" /db_xref="SGD:S0002324" 390 a 226 c 240 g 352 t			
BASE COUNT	390 a 226 c 240 g 352 t			
ORIGIN				
Query Match	79.0%; Score 15.8; DB 7; Length 1208;			
Best Local Similarity	89.5%; Pred. No. 73;			
Matches	17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			

/codon_start=1
/db_xref="SGD:S0002324"
/protein_id="CAA27005.1"
/db_xref="PID:g3515"
/db_xref="GI:3315"
/db_xref="SWISS-PROT:P04819"
/translation="MRRLLTGCLLSARPLKSRLLPMLSSLPSSAGKKPKQATLARE
FTSMKNKTEGTPSPKSKHMLEERMDNVGEEYATKKLQTAVTHTVAAPSSMG
NFISSPSSAPSGVADSPQOSQRLVGEVDALSSNNNDHYSNNIPYSEVCEVFNKIEA
ISSRLFIIRICSDFIKIMKQSKNVIPITYLFINRLGPDYEAGLIGENLLAKTI
STCGKSKSOIKLKYKDGLGELIANGARNVQTFKPKPLTVGEVFNRLRAIKTQG
KDSQLKMLIRKLTACKIGIAKFLIRLSLEKRLIGLAEKTVLISLSKALLHDNEN
EDSPKDPMDVESAQQIRDAFCQVNYEIVINSLGHEIMNLDKTYTLRPGPLK
PMLAKPTAINVLFVDFGETFTSEYKIDGERAQVHLLNDGMTMRIYSRNGENMTIRY
EINIDFIOOLDTNNKILDCAVAMDQKGLPFOVLSTRKRKDVDELNDVKVKCL
FADFDLCYNDERLNLKSLAKEREYLTKVTKVVPGEQYATQITNNLDELQFDES
NHSCEGLAVKMLEGPESHYEPKSRNWLKKDYLEGVGDSDLCVLGAYGKGRKT
GTYGFLGCLNQDTEGFTCKGIGTGFSDMLQLLHDRLTPTIIDGPKATVFDSSA
EPDVMFETTLFEVLTAADLSPIYKAGSATFDKGVSLRFPRLRDKGVEDATSS
DQIVELYENQSHMQN"
complement(3101..3676)
/gene="CDC36"
/note="open reading frame (CDC36) (aa 1-191)"
/codon_start=1
/db_xref="SGD:S0002324"
/protein_id="CAA27006.1"
/db_xref="PID:g3516"
/db_xref="GI:3316"
/db_xref="SWISS-PROT:P06100"
/translation="MEKFGKALVPLKKLEDKELSSYDHSMTLGADLSMLYSLGIP
RDSQHRVLDTFQSPAETSSEVEPRFTFPTNIPGVLSQSTVTPPCFNSIQNDQ
RVAFQDETLFLFYKHPGTVIQELTYLELRKNRWYHKTLLKAWLTQDPMMEPIVSAD
GLSERGSYVFEDPQWEKQCORDFLLFYNAIM"
3135..3137
/gene="CDC36"
/note="minor 3' polyadenylation site"
/db_xref="SGD:S0002324"
3142..3144
/gene="CDC36"
/note="major 3' polyadenylation site"
/db_xref="SGD:S0002324"
1165 a 755 c 771 g 1158 t

BASE COUNT 1165 a 755 c 771 g 1158 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 7; Length 3849;
Best Local Similarity 89.5%; Pred. No. 77;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACATATGACCTTCCATG 20
||||||| | ||||| |
Db 3611 TACATATGACCATCCATG 3593

RESULT 14
SCYDL165W 944 bp DNA PLN 11-AUG-1997
LOCUS
DEFINITION S.cerevisiae chromosome IV reading frame ORF YDL165W.
ACCESSION 74213 Z71256
NID 91431261
VERSION 274213.1 GI:1431261
KEYWORDS baker's yeast.
SOURCE
ORGANISM
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 944)
AUTHORS Pohl,T.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 944)
AUTHORS MIPS.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-1996) Data collected by MIPS on behalf of the

European yeast chromosome IV sequencing project. MIPS at the
Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152
Martinsried, FRG; E-mail: Mewes@mips.emblnet.org

FEATURES
Source
Location/Qualifiers
1..944
/organism="Saccharomyces cerevisiae"
/db_xref="taxon:4932"
/chromosome="IV"
248..823
/gene="CDC36"
/db_xref="SGD:S0002324"
248..823
/gene="CDC36"
/note="ORF YDL165W"
/codon_start=1
/db_xref="SGD:S0002324"
/protein_id="CAA98739.1"
/db_xref="PID:g253074"
/db_xref="PID:g1431262"
/db_xref="GI:1431262"
/db_xref="SWISS-PROT:P06100"
/translation="MEKFGKALVPLKKLEDKELSSYDHSMTLGADLSMLYSLGIP
RDSQHRVLDTFQSPAETSSEVEPRFTFPTNIPGVLSQSTVTPPCFNSIQNDQ
RVAFQDETLFLFYKHPGTVIQELTYLELRKNRWYHKTLLKAWLTQDPMMEPIVSAD
GLSERGSYVFEDPQWEKQCORDFLLFYNAIM"
315 a 171 c 178 g 280 t

BASE COUNT 315 a 171 c 178 g 280 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 7; Length 944;
Best Local Similarity 89.5%; Pred. No. 72;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACATATGACCTTCCATG 20
||||||| | ||||| |
Db 313 TACATATGACCATCCATG 331

RESULT 15
SCSFAARP 7008 bp DNA PLN 29-NOV-1994
LOCUS
DEFINITION S.cerevisiae SFA and ARP genes.
ACCESSION X68020 S59849
NID 9577609
VERSION X68020.1 GI:577609
KEYWORDS arp gene; beta-D-galactosidase; CDC36 gene; CDC9 gene; sfa gene.
SOURCE baker's yeast.
ORGANISM
Saccharomyces cerevisiae
Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 7008)
AUTHORS Wehner,E.P., Rao,E. and Brendel,M.
TITLE Molecular structure and genetic regulation of SFA, a gene
responsible for resistance to formaldehyde in Saccharomyces
cerevisiae, and characterization of its protein product
JOURNAL Mol. Gen. Genet. 237 (3), 351-358 (1993)
MEDLINE 93247548
REFERENCE 2 (bases 1 to 7008)
AUTHORS Wehner,E.
TITLE Direct Submission
JOURNAL Submitted (05-AUG-1992) E. Wehner, Institut f Mikrobiologie,
Theodor Steinkai 7, Haus 75, 6000 Frankfurt/Main, FRG
REFERENCE 3 (bases 1 to 7008)
AUTHORS Wehner,E.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-1994) E. Wehner, Institut f Mikrobiologie,
Theodor Steinkai 7, Haus 75, 6000 Frankfurt/Main, FRG
COMMENT On Nov 28, 1994 this sequence version replaced gi:288588.
FEATURES
Location/Qualifiers
1..7008
/organism="Saccharomyces cerevisiae"
/strain="AH22"
/db_xref="taxon:4932"

gene
1. .442
/gene="CDC9"
CDS
<1. .442
/gene="CDC9"
/codon_start=2
/protein_id="CAA48158.1"
/db_xref="PID:g600039"
/db_xref="GI:600039"
/db_xref="SWISS-PROT:P04819"

/translation="DLGVLYGRGKRTGYGGFLGCGYNODTGEFETCKIGTGFS
DEMLOLHDLRTIIDEPKATFVDSASPDVWFETTLFEVLTABLESLSPIYKAGS
ATFDKGVSLRFPFLRLREDKGVEDATSSQIVELYENQSHMQN"
complement(565. .1140)
/gene="CDC36"
/db_xref="SGD:S0002324"
complement(565. .1140)
/gene="CDC36"

gene
/codon_start=1
/db_xref="SGD:S0002324"
CDS
/protein_id="CAA48160.1"
/db_xref="PID:g577610"
/db_xref="GI:577610"
/db_xref="SWISS-PROT:P06100"

/translation="MERFGLKALVPLKLEDKELSSYDHSMTLGADLSSMLYSLGIP
RDSQDHVLDTFQSPWAEATSRSEVEPRFTPESTINIPGVLSQSTVPPCFNSIQNDQO
RLVLFQDETFFLFYKHPTGVIQELTYLELRKRNRYHKTWKAWLTAKDPWMEPIVSAD
GLSERGSYVFEDPQWEKQCQDFLLFYNAIM"
2277. .4436
/gene="ARP"

gene
/codon_start=1
CDS
/db_xref="SGD:S0002326"
2277. .4436
/gene="ARP"

gene
/codon_start=1
CDS
/db_xref="SGD:S0002326"
/protein_id="CAA48159.1"
/db_xref="PID:g288590"
/db_xref="GI:288590"

/db_xref="SWISS-PROT:P32770"
/translation="MHYVLELOVALPDTPKQCRANTAFQIVNAETLVCHVGTNS
LPSIEVNGTKVLESAMVQDKDIHDVIGNDDEFVLSLYSTWHIRVTLPRQARDGFI
LTSYLQHPKVFDELKEDFCVNHPEILGOKKALSNNCNTKSSINAARNTKDLDEI
VRILEVSPIEEAGSVPEIYSLKRTDILQHKKCTSPEDMESVLTKPYSHTDIR
AFLOESKILLYMNLPPDTQSELESFTQYGVPGVFTVKNIQVEDTSNNNNWSLN
NSPYVEDISIGFVVFQTHEEATEVIALNGRSLSLNANTKOPRVVEHYLELQPSST
GVLDKAOEILSPFPQSKNPGDWCPCSGSFNFORRTACFRCSPAPSPSOIHTAN
SNNVNRRNLNARNVSGSSNISNTAANHYCAPFENMIANTPAALTYNRAHPA
ITPLSRQSLNMAPSRSGSPIIADHESGNNNIAPYRYNNNNNNNNNNNNNNNNRY
NINNINGNGNGNNNNNNNNNNNNNNNNHNSINSNTNNNNNNNNNNNNNNNN
SNIGMGCGSNMPFRAGDWKCSTCTYHNFKNYVCLRCGGPKSISGDASETNHYIDSS
TFGPASRTPSNNNISVNTNGSGNAGRTDGNDRGDISLMEFMSPLSMATKSMKEGD
NGSSSFNEFKSKANVNFSTVUNSAFNGNGNSIRW"
complement(4667. .582 /)
/gene="SFA"

gene
/db_xref="SGD:S0002327"
CDS
complement(4667. .582 /)
/gene="SFA"

gene
/codon_start=1
CDS
/db_xref="SGD:S0002327"
/protein_id="CAA48161.1"
/db_xref="PID:g288591"
/db_xref="GI:288591"

/db_xref="SWISS-PROT:P32771"
/translation="MSAATVGPKICIAAVAYDAKKPLSVEEITVDAPKAHEVRKIE
YTAVCHTDAYTLGSDPEGLPCVLGHEGAGIVESYCDVDVITVKPGDHVIALYAECC
KRCFTSGKTNLCGAVRATOGKGVMPDGTTRFNKAGEDIYHEMGSTFSEIYTVADY
SVVAIDPAPLDAACLLCGVTTGFGAALATANVQKGDVAVFCGTVGLSVIOGAKL
RGASKIADIDNNKKOYCSQFATDVFNPKEDLAKDQITIVERKLIEMTDGGLDFTDC
TGNTKIMRDALAEACHRGQSQSIIGVAAAGEEISTRPVLTVGRVWKGSAFGGIKGRS
EMGLIKDYQKALKVVEEFTTHRRPFKEINQAFEDLHNGDCLRTVLKSDIEK"

BASE COUNT 2266 a 1305 c 1367 g 2070 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 8; Length 7008;
Best Local Similarity 89.5%; Pred. No. 79;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TACATATGAGCCTTCCATG 20
||||||| |
Db 1075 TACATATGACCATTCATG 1057

Search completed: September 18, 1999, 15:49:29
Job time: 1449 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 05:27:24 ; Search time 213.04 Seconds
(without alignments)
23.488 Million cell updates/sec

Title: US-09-037-472-8
Perfect score: 20
Sequence: 1 TTACATATGAGCCTTCATG 20

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	1	Primer for detecti
2	20	100.0	20	1	Interleukin-1-alpha
3	20	100.0	20	1	Interleukin 1 (441
c 4	18.4	92.0	5701	1	Human IL-1ra BAC c
5	16.4	82.0	6265	1	Urea amidolyase ge
c 6	15.8	79.0	4708	1	Human acetyl choli
7	15.8	79.0	4708	1	Human acetyl choli
8	15.8	79.0	7158	1	HEV strain SAR-55
9	15.8	79.0	7158	1	Hepatitis E virus
c 10	15.2	76.0	354	1	Expressed Sequence
c 11	15.2	76.0	354	1	Human brain Expre
c 12	15.2	76.0	8148	1	Streptococcus pneu
13	15.2	76.0	36	1	PCR primer used to
14	14.8	74.0	1891	1	Potato citrate syn
15	14.8	74.0	1747	1	Potato citrate syn
16	14.8	74.0	1891	1	Potato citrate syn
17	14.8	74.0	110000	1	Continuation (2 of
18	14.8	74.0	110000	1	Continuation (2 of
19	14.2	71.0	7833	1	Human papilloma vi
20	14.2	71.0	2387	1	Brn-1 cDNA. Polype
21	14.2	71.0	8841	1	Tryptophan-2,3-dio
22	14.2	71.0	4782	1	Rat homologue of h
23	14.2	71.0	33	1	Human immunodefici
24	14.2	71.0	2387	1	Wasp Brh-1 toxin g
25	14.2	71.0	467	1	Human papillomavir
26	14.2	71.0	288	1	DNA sequence of L1
27	14.2	71.0	467	1	DNA encoding human
c 28	14.2	71.0	2846	1	Human TRAF5 cDNA.
c 29	14.2	71.0	3993	1	Human papilloma vi
30	14.2	71.0	455	1	Rat kidney cell ca
c 31	14.2	71.0	4131	1	HIV-1 proviral DNA
c 32	14.2	71.0	33	1	Rat kidney calcium
c 33	14.2	71.0	4131	1	Target sequence in
34	14.2	71.0	145	1	Erlichia canis VS
c 35	14.2	71.0	1570	1	Candida neoformans
36	14.2	71.0	3168	1	Candida neoformans
37	14.2	71.0	3220	1	Candida neoformans
38	14.2	71.0	3168	1	Candida neoformans
39	14.2	71.0	610	1	Staphylococcus aur
c 40	14.2	71.0	4131	1	Rat parathyroid ca
c 41	14.2	71.0	4718	1	Mouse semaphorin r
c 42	14.2	71.0	4733	1	Mouse semaphorin r
c 43	14.2	71.0	4769	1	Mouse semaphorin r

Mouse semaphorin r
DNA encoding p30a

c 44 14.2 71.0 4784 1 X09000
c 45 14.2 71.0 864 1 X34760

ALIGNMENTS

```

RESULT 1
Tl3884
ID Tl3884 standard; DNA; 20 BP.
AC Tl3884;
DT 19-OCT-1997 (first entry)
DE Primer for detecting genetic predisposition to periodontal disease.
KW Periodontal disease; gingivitis; periodontitis; polymorphism;
KW Interleukin-1 alpha; IL-1A; interleukin-1 beta; IL-1B; primer; PCR;
KW polymerase chain reaction; ss.
OS Synthetic.
PN WO9706180-A1.
PD 20-FEB-1997.
PF 02-AUG-1996; U12455.
PR 03-AUG-1995; US-510696.
PA (KORN/) KORNMAN K S.
PA (MED/) MEDICAL SCI SYSTEMS INC.
PI Duff GW, Kornman KS;
DR WPI: 97-154207/14.
PT Identification of patient's genetic polymorphism pattern - allows
PT prediction of increased periodontal disease severity
PS Claim 3; Page 25; 33pp; English.
CC PCR primers (T70316 and Tl3884-88) can be used in a new method
CC for identifying a genetic predisposition to periodontal diseases by
CC detecting the presence of DNA polymorphisms in the gene sequences
CC for interleukin-1 alpha (IL-1A) and interleukin-1 beta (IL-1B).
CC Alleles associated with severe disease were identified as IL-1A
CC allele 2 together with IL-1B (TagI) allele 2. The primers are used
CC to amplify DNA from a blood or tissue sample and products are
CC subjected to restriction digestion to determine the polymorphism
CC pattern of the patient. A single base variation (C/T) polymorphism
CC at IL-1A base -889 can be identified using primers (T70316 and
CC Tl3884) corresponding to bases -967 to -945 and -888 to -869,
CC respectively. An NcoI site is created if C is available at -889,
CC but not if T is present.
SQ Sequence 20 BP; 5 A; 5 C; 3 G; 7 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.062;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCATG 20
   |||||
DB 1 TTACATATGAGCCTTCATG 20

RESULT 2
V32396
ID V32396 standard; DNA; 20 BP.
AC V32396;
DT 11-SEP-1998 (first entry)
DE Interleukin-1-alpha primer 4 (-888/-869).
KW IL-1-alpha; genetic polymorphism; PCR; primer; amplification;
KW sight threatening diabetic retinopathy; interleukin-1-alpha;
KW interleukin-1-beta; interleukin-1RN; ss.
OS Synthetic.
PN WO9815653-A1.
PD 16-APR-1998.
PF 09-OCT-1997; G02790.
PR 10-OCT-1996; GB-021129.
PA (DUFF/) DUFF G.
PA (RENN/) RENNIE I.
PA (RICH/) RICHARDSON R.
PI Duff G, Rennie I, Richardson R;
DR WPI: 98-240835/21.

```

PT Predicting increased risk of sight-threatening diabetic retinopathy
 PT - comprises identifying genetic polymorphism pattern for genes
 PT IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
 PT symptoms occur
 PS Claim 2; Page 33; 41pp; English.
 CC Interleukin-1-alpha (IL-1-alpha) primers 4 and 3 (V32395) were used to
 CC amplify the IL-1-alpha gene region to identify single base variation
 CC polymorphism of C/T at base 889. The invention claims to provide a
 CC method for predicting the risk of sight threatening diabetic retinopathy.
 CC The method involves isolating DNA from a patient and determining the DNA
 CC polymorphism pattern of the genes that code for interleukin-1-alpha,
 CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
 CC identified is then compared with controls of known DNA polymorphism
 CC patterns thereby identifying patients carrying a genetic polymorphism
 CC associated with increased risk of sight threatening diabetic retinopathy.
 CC The method may be able to identify diabetic patients at risk before the
 CC clinically detectable disorders occur. Polymorphism pattern
 CC determination of IL genes involved PCR reactions using primers V32389-
 CC V32398. The method is also claimed to be useful in conjunction with
 CC identification of other genes associated with sight threatening diabetic
 CC retinopathy in genomic DNA and therefore, in identifying diabetic
 CC patients expressing multiple risk patterns.
 SQ Sequence 20 BP: 5 A; 5 C; 3 G; 7 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.062;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
 |||||

Db 1 TTACATATGAGCCTTCCATG 20

RESULT 3

X16612
 ID X16612 standard; DNA: 20 BP.

AC X16612;

DT 29-APR-1999 (first entry)

DE Interleukin 1 (44112332) haplotype PCR primer #6.

KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;

KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;

KW Graves disease; systemic lupus erythematosus; lichen sclerosis;

KW ulcerative colitis; PCR primer; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9854359-A1.

PD 03-DEC-1998.

PF 21-MAY-1998; G01481.

PR 29-MAY-1997; GB-011040.

PA (CAMP//) CAMP N J.

PA (COXA//) COX A.

PA (DGIO//) DE GIOVINE F S.

PA (DUFF//) DUFF G.

PI Camp NJ, Cox A, De Giovine FS, Duff G;

PT WPI: 99-080814/07.

PT New method of determining a patient's susceptibility to inflammatory

PT disorders - by detecting the presence of an IL-1 (44112332)

PT haplotype, useful in designing treatment strategies that modulate

PT the activity of proteins produced by the IL-1 gene cluster

PS Claim 3; Page 33; 49pp; English.

CC A method has been developed for determining a patient's susceptibility

CC to an inflammatory disorder. The method comprises the detection of an

CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the

CC patient, where its presence indicates susceptibility to an inflammatory

CC disorder. X16607 to X16631 represent PCR primer used in the method for

CC detecting the IL-1 (44112332) haplotype. The method provides kits for

CC the early prediction of a patient's susceptibility to inflammatory

CC disorders, including coronary artery disease, osteoporosis, nephropathy

CC in diabetes mellitus, alopecia areata, Graves disease, systemic lupus

CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of

CC alleles of the haplotype can be applied to particular inflammatory

CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,

CC

CC

CC

CC

CC

CC

CC

CC psoriasis, and insulin dependent diabetes. The identification of a
 CC disease-associated haplotype enables determination of which alleles are
 CC causative, and this information is useful in designing treatment
 CC strategies, including gene therapy and treatment using various agents
 CC that modulate the activity of proteins produced by the IL-1 gene cluster.
 CC Some alleles from the IL-1 gene cluster are associated with particular
 CC inflammatory diseases, and insufficient IL-1 production appears to act
 CC centrally in the pathology of these diseases. Therefore, the use of IL-1
 CC gene clusters is useful in determining genetic susceptibility to
 CC inflammatory diseases, including those with a multifactorial etiology
 CC with a polygenic component.
 SQ Sequence 20 BP: 5 A; 5 C; 3 G; 7 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.062;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
 |||||

Db 1 TTACATATGAGCCTTCCATG 20

RESULT 4

X02988/c

ID X02988 standard; DNA: 5701 BP.

AC X02988;

DT 22-JUN-1999 (first entry)

DE Human IL-1ra BAC contiguous DNA sequence 33.

KW TANGO-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;

KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;

KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;

KW growth factors; treatment; IL-1 receptor complex; BAC; ss.

OS Homo sapiens.

PN WO9906426-A1.

PD 11-FEB-1999.

PF 03-AUG-1998; U16102.

PR 02-JUL-1998; US-091650.

PR 04-AUG-1997; US-054646.

PA (MLL-) MILLENNIUM BIOTHERAPEUTICS INC.

PI Pan Y;

PT WPI: 99-153692/13.

PT New isolated nucleic acid encoding the new human cytokine TANGO-77 -

PT used to inhibit inflammation and to screen for specific modulators

PS Example 5; Figure 3; 226pp; English.

CC X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences

CC containing alternatively spliced forms of human IL-1ra. Such fragments

CC are used in the method of the invention which describes the isolation of

CC a novel human TANGO-77 encoding nucleic acid and protein. TANGO-77 is a

CC member of the cytokine superfamily that is expected to inhibit

CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may

CC also bind to a new receptor so could regulate other cellular processes

CC associated with acute or chronic inflammation, e.g. asthma, chronic

CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory

CC bowel disease. It may also induce or suppress interleukins, cytokines and

CC growth factors. Modulators of this protein are used to treat or prevent

CC conditions associated with abnormal levels of inflammation, or activity

CC of IL-1 or its receptor complex.

CC Sequence 5701 BP: 1729 A; 1238 C; 1079 G; 1654 T;

Query Match 92.0%; Score 18.4; DB 1; Length 5701;
 Best Local Similarity 95.0%; Pred. No. 0.81;
 Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
 |||||

Db 529 TTACATATGAGCCTTCCATG 510

RESULT 5

Q49460

ID Q49460 standard cDNA; 6265 BP.

```

AC Q49460:
DE 10-MAY-1994 (first entry)
DE Urea amidolyase gene.
KW Urea amidolyase: URL; yeast; recombinant plasmid; ss.
OS Saccharomyces cerevisiae.
FH Key Location/Qualifiers
FT cds 627..6134
FT /tag= a
FT /product= URL
FT J05244959-A.
PN 24-SEP-1993.
PD 05-MAR-1992; 084531.
PF 05-MAR-1992; JP-084531.
PR (TOYM ) TOYOBOKK.
PA WPI: 93-338925/43.
DR P-PSDB; R42839.
PT DNA having the genetic information of urea amidolyase originated
PT from Saccharomyces yeast - can be used to prepare high purity
PT urea amidolyase by culturing the transformant comprising the DNA
PS Claim 1; Page 10-17; 17pp; Japanese.
CC This sequence encodes a protein which has urea amidolyase (URL)-
CC activity and is derived from yeast. This sequence may be used within
CC a recombinant plasmid for the production of highly pure URL.
SQ Sequence 6265 BP; 1873 A; 1156 C; 1361 G; 1875 T;

Query Match 82.0%; Score 16.4; DB 1; Length 6265;
Best Local Similarity 94.4%; Pred. No. 9.2; Indels 0; Gaps 0;
Matches 17; Conservative 0; Mismatches 1;

QY 1 TTACATATGAGCCTTCCA 18
|| |||||
DB 937 TTGCATATGAGCCTTCCA 954

RESULT 6
ID N50415 standard; DNA; 4708 BP.
AC N50415;
DE 08-JAN-1992 (first entry)
DE Human acetyl choline receptor alpha gene.
KW hAcr-alpha; myasthenia gravis; MG; neuromyopathy; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 46..3126
FT /tag= a
FT mat_peptide 346..3123
FT /tag= b
FT intron 89..328
FT /tag= c
FT /note= "60N inserted as a spacer"
FT intron 475..585
FT /tag= d
FT intron 631..917
FT /tag= e
FT /note= "60N inserted as spacer"
FT intron 1028..1200
FT /tag= f
FT /note= "60N inserted as spacer"
FT intron 1397..1624
FT /tag= g
FT /note= "60N inserted as spacer"
FT intron 1863..2051
FT /tag= h
FT /note= "60N inserted as spacer"
FT intron 2276..2430
FT /tag= i
FT /note= "60N inserted as spacer"
FT intron 2671..2994
FT /tag= j
PN J60078996-A.
PD 04-MAY-1985.
PF 05-OCT-1983; 186402.

```

```

PR 05-OCT-1983; JP-186402.
PA (MITU ) MITSUBISHI CHEM IND KK.
DR WPI: 85-144120/24.
DR P-PSDB; P50361.
PT New peptide(s) useful in treatment of myasthenia gravis - also in
PT diagnosis, having acetylcholine receptor-alpha like activities.
PS Disclosure; Fig 1; 13pp; Japanese.
CC Peptides derived from the hAcr-alpha gene product have similar
CC immunogenicity to the gene product, and may be useful in the
CC treatment and diagnosis of myasthenia gravis caused by neuropathy in
CC autoimmune reaction to ACR.
CC The sequence given contains breaks in some of the introns, at these
CC locations 60 N residues have been inserted to preserve integrity.
SQ Sequence 4708 BP; 1124 A; 1059 C; 965 G; 1200 T;

Query Match 79.0%; Score 15.8; DB 1; Length 4708;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCAT 19
|| |||||
DB 1164 TGACAGATGAGCCTTCCAT 1146

RESULT 7
Q45197
ID Q45197 standard; cDNA; 7168 BP.
AC Q45197;
DE 21-OCT-1994 (first entry)
DE HEV strain SAR-55 cDNA sequence.
KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
KW antibody; detection; diagnosis; primates; stool suspension; ss.
OS Hepatitis E virus strain SAR-55.
FH Key Location/Qualifiers
FT cds 28..5109
FT /tag= a
FT /label= ORF-1
FT misc_difference 3739..3741
FT /tag= b
FT /codon= seq:cag, aa:Glu
FT misc_difference 3757..3759
FT /tag= c
FT /codon= seq:cag, aa:Glu
FT misc_difference 4081..4083
FT /tag= d
FT /codon= seq:gtg, aa:Glu
FT misc_difference 5011..5013
FT /tag= e
FT /codon= seq:ggc, aa:Glu
FT cds 5147..7129
FT /tag= f
FT /label= ORF-2
FT misc_difference 5780..5782
FT /tag= g
FT /codon= seq:tgg, aa:Tyr
FT cds 5106..5457
FT /tag= h
FT /label= ORF-3
PN W09406913-A.
PD 31-MAR-1994.
PF 17-SEP-1993; U08849.
PR 18-SEP-1992; US-947263.
PA (USSH ) US SEC DEPT HEALTH.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI: 94-118462/14.
DR P-PSDB; R51264-66.
PT Purified hepatitis E strain SAR-55 virus - used to develop prods.
PT for use in detection, diagnosis, vaccines and therapy of
PT hepatitis E virus infection
PS Claim 2; Page 16-20; 114pp; English.
CC This sequence represents the genomic sequence of the hepatitis E
CC virus (HEV) strain SAR-55. This sequence contains three open reading

```


CC frames (ORFs). The proteins encoded by this sequence can be used to
CC stimulate the production of protective antibodies upon injection into
CC a mammal that would serve to protect the mammal upon challenge with
CC wild type HEV. The proteins can be used for detection and diagnosis
CC of HEV infection. This cDNA was isolated from primates inoculated
CC with stool suspensions obtained from hepatitis E patients.
SQ Sequence 7168 BP; 1223 A; 2294 C; 1867 G; 1784 T;

Query Match 79.0%; Score 15.8; DB 1; Length 7168;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCAT 19
||||| ||||||| ||
Db 2189 TTACATCTGAGCCTTCTAT 2207

RESULT 8
ID T27394 standard; cDNA; 7158 BP.
AC T27394;
DT 26-NOV-1996 (first entry)
DE Hepatitis E virus strain SAR-55 cDNA (ATCC 75302).
KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
KW Structural region; antigen; detection; antibody; vaccine;
KW Immunisation; infection; ss.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT 28..5099
FT /*tag= a
FT /label= ORF-1 (R91813)
FT /transl_except= pos:3739..3741, aa:Glu
FT /note= "10 bp nucleic acid sequence TGCTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 for numbering to conform to
FT that given in the specification"
FT 5137..7119
FT /*tag= b
FT /label= ORF-2 (R91814)
FT /note= "10 bp nucleic acid sequence TGCTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 for numbering to conform to
FT that given in the specification"
FT 5096..5467
FT /*tag= c
FT /label= ORF-3 (R91815)
FT /note= "10 bp nucleic acid sequence TGCTNTTYGA
FT has to be inserted between nucleotides
FT 4390..4391 for numbering to conform to
FT that given in the specification"
PN WO9610580-A2.
PD 11-APR-1996.
PE 03-OCT-1995; U13102.
PR 03-OCT-1994; US-316765.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Tsarev SA;
DR WPI; 96-209320/21.
DR P-PSDB; R91813, R91814, R91815.
PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT antigenic protein useful in diagnosis, prophylaxis and treatment of
PT hepatitis E virus infection
PS Claim 2; Pages 16-21; 121pp; English.
CC The present sequence is the cDNA of the hepatitis E virus (HEV)
CC strain SAR-55, which was implicated in an enterically transmitted
CC non-A, non-B hepatitis in Pakistan. The protein encoded by the
CC structural region of the virus (i.e. ORF-2), which is capable of
CC forming HEV like particles, is useful for the detection of HEV
CC antibodies (pref. IgG or IgM) in blood, plasma, sera,
CC cerebrospinal fluid, tissue, urine or pleural fluid. The protein,
CC and anti-HEV antibodies generated using the protein, can also be
CC used in vaccines for immunising an animal against HEV infection.
CC The protein is identified as a band of greater than 50 kD

CC following SDS-PAGE of cell lysates of insect cells infected with
CC a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
CC expression vectors pPIC9-1779, -1780 and -1781.
SQ Sequence 7158 BP; 1221 A; 2293 C; 1864 G; 1780 T;

Query Match 79.0%; Score 15.8; DB 1; Length 7158;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCAT 19
||||| ||||||| ||
Db 2189 TTACATCTGAGCCTTCTAT 2207

RESULT 9
ID V71604 standard; DNA; 7168 BP.
AC V71604;
DT 02-FEB-1999 (first entry)
DE Hepatitis E virus (HEV) polypeptides encoding nucleic acid SAR-55.
KW Hepatitis E virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
KW Passive immunisation; ss.
OS Hepatitis E virus.
FH Key Location/Qualifiers
FT 28..5109
FT /*tag= a
FT /transl_except= (pos:3739..3741, aa:Glu)
FT /transl_except= (pos:3757..3759, aa:Glu)
FT /transl_except= (pos:4081..4083, aa:Glu)
FT /transl_except= (pos:5011..5013, aa:Glu)
FT /product= "ORF-1 protein"
FT 5147..7129
FT /*tag= b
FT /transl_except= (pos:5780..5782, aa:Tyr)
FT /product= "ORF-2 protein"
FT 5106..5477
FT /*tag= c
FT /product= "ORF-3 protein"
PN WO9846761-A1.
PD 22-OCT-1998.
PE 09-APR-1998; U07418.
PR 11-APR-1997; US-840316.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PI Emerson SU, Purcell RH, Robinson RA, Tsarev SA;
DR WPI; 98-568733/48.
DR P-PSDB; W81519, W81520, W81521.
PT New hepatitis E virus DNA from Pakistani strain SAR-55 - used for,
PT e.g. developing products for diagnosis of, and vaccination against
PT hepatitis E virus infection
PS Disclosure; Pages 126-131; 204pp; English.
CC This represents a DNA sequence designated SAR-55 encoding hepatitis E
CC virus (HEV) ORF proteins ORF-1, ORF-2 and ORF-3. A host organism
CC transformed or transfected with a recombinant expression vector
CC containing the SAR-55 nucleic acid can be used to produce the HEV
CC proteins, especially ORF-2 protein. The recombinant HEV proteins can be
CC used as diagnostic agents and as vaccines for use against HEV infection.
CC The detection of antibodies specific for HEV can be used for the
CC diagnosis of infection and diseases caused by HEV, and for monitoring the
CC progression of such disease. Such methods are also useful for monitoring
CC the efficacy of therapeutic agents during the course of treatment of HEV
CC infection and disease in a mammal. The antibodies can be used for
CC detection or for passive immunisation of mammals.
SQ Sequence 7168 BP; 1222 A; 2294 C; 1868 G; 1784 T;

Query Match 79.0%; Score 15.8; DB 1; Length 7168;
Best Local Similarity 89.5%; Pred. No. 19;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCAT 19
||||| ||||||| ||
Db 2189 TTACATCTGAGCCTTCTAT 2207

CC coding-region prediction program CRM. See also Q59041-Q61440.
SQ Sequence 354 BP; 123 A; 61 C; 92 G; 70 T;

Query Match 76.0%; Score 15.2; DB 1; Length 354;
Best Local Similarity 85.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
Db 62 TTTCATGAGCCTTCCATG 43

RESULT 12
V52144/C

ID V52144 standard; DNA; 8148 BP.

AC V52144;

DE 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:11.

KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

OS Streptococcus pneumoniae.

PN W09818931-A2.

PD 07-MAY-1998.

PF 30-OCT-1997; U19588.

PR 31-OCT-1996; US-029960.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,

PI Kunsch CA, Rosen CA;

DR WPI; 98-27225/24.

PT Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

PS Claim 1; Page 220-224; 1409pp; English.

CC The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded

CC on it, or a representative fragment or a sequence at least 95% identical

CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1

CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus

CC pneumoniae. The present invention also describes an isolated nucleic acid

CC molecule encoding a homologue of any of the fragments of the S.pneumoniae

CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced

CC by a process comprising: (a) screening a genomic DNA library using as a

CC probe a target sequence defined by any of the sequences in SEQ ID NO:1

CC to 391, identifying members of the library which contain sequences

CC that hybridize to the target sequence and isolating the nucleic acid

CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced

CC from an organism, amplifying nucleic acid molecules whose nucleotide

CC sequence is homologous to amplification primers derived from the

CC fragment of the S. pneumoniae genome to prime the amplification and

CC isolating the amplified sequences. The computer readable medium can be

CC used in a computer-based system for identifying fragments of the

CC S. pneumoniae genome of commercial importance, or expression modulating

CC fragments of the S. pneumoniae genome. Products from the present

CC invention can be used in diagnosis kits and assays, and pharmaceutical

CC compositions and vaccines for S. pneumoniae.

SQ Sequence 8148 BP; 2475 A; 1445 C; 1863 G; 2365 T;

Query Match 76.0%; Score 15.2; DB 1; Length 8148;
Best Local Similarity 85.0%; Pred. No. 41;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
Db 5301 TTACATAGGATCTTCCATG 5282

RESULT 13
V42902

ID V42902 standard; DNA; 36 BP.

AC V42902;

RESULT 10
Q39643/C

ID Q39643 standard; DNA; 354 BP.

AC Q39643;

DE 20-MAY-1993 (first entry)

DE Expressed Sequence Tag human gene marker EST00020.

KW human gene sequencing; PCR mapping; somatic cell hybrids;

KW sublocalisation; gene tagging; tissue typing.

OS Synthetic.

PN W09300353-A.

PD 07-JAN-1993.

PF 19-JUN-1992; U05222.

PR 20-JUN-1991; US-716831.

PR 12-FEB-1992; US-837195.

PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

PI Adams MD, Venter CJ;

DR WPI; 93-036325/04.

PT Particular expressed sequence tags from human CDNA - corresponds

PT to transcription prods. of genes, useful for tagging genes,

PT mapping chromosomes and tissue typing

PS Claim 3; Page 86; 199pp; English.

CC This sequence represents an EST (expressed sequence tag) ESTs are markers

CC for human genes actually transcribed in vivo. Unlike the random genomic

CC DNA sequence tagged sites (STSs), ESTs point directly to expressed genes.

CC The use of ESTs could facilitate the tagging of most expressed human

CC genes within a few years at a fraction of the cost of complete genomic

CC sequencing. Using PCR primers Q39419-Q39580 (sequences designed

CC from the ESTs) sublocalisation of an EST can be achieved with panels of

CC fragments from specific chromosomes or pools of large genomic clones in

CC an analogous manner. This sequence represents EST00020.

SQ Sequence 354 BP; 126 A; 62 C; 90 G; 70 T;

Query Match 76.0%; Score 15.2; DB 1; Length 354;
Best Local Similarity 85.0%; Pred. No. 29;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
Db 62 TTTCATGAGCCTTCCATG 43

RESULT 11
Q59055/C

ID Q59055 standard; CDNA; 354 BP.

AC Q59055;

DE 16-MAR-1994 (first entry)

DE Human brain Expressed Sequence Tag EST00020.

KW Gene transcription product; genetic markers; tagging; in vivo;

KW transcription; mapping; locations; chromosomes; chromosomal; ss.

OS Homo sapiens.

PN W09316178-A.

PD 19-AUG-1993.

PF 12-FEB-1993; U01294.

PR 12-FEB-1992; US-837195.

PA (USSH) US DEPT HEALTH & HUMAN SERVICE.

PI Adams MD, Moreno RF, Venter CJ;

DR WPI; 93-272882/34.

PT Enriched oligonucleotides and corresp. sequences - used as

PT markers for human genes transcribed in-vivo, facilitate tagging

PT of most human genes

PS Example 1; Page 114; 500pp; English.

CC The Expressed Sequence Tag was isolated from a human brain cDNA

CC library as part of a large set of ESTs which can be used as markers

CC for human genes transcribed in vivo. They can be used to facilitate

CC tagging of most human genes, for mapping locations of expressed genes

CC on chromosomes, for individual or forensic identification, for mapping

CC locations of disease-associated genes, for identification of tissue

CC type, and for prepn. of antisense sequences, probes and constructs.

CC EST00020 has an "excellent" coding probability as evaluated using the

DT 22-OCT-1998 (first entry)
 DE PCR primer used to mutate part of genome of adenovirus serotype 5.
 KW Domain alpha 2; antigen; major histocompatibility complex class I;
 KW MHC-I; primary receptor; serotype C adenovirus; human; fibronectin;
 KW module III; targeting; cell-surface protein; cellular receptor;
 KW bifunctional ligand; tumour cell; infected cell; PCR primer; ss.
 OS Synthetic.
 OS Mastadenovirus.
 PN FR2758822-AL.
 PD 31-JUL-1998.
 PF 09-SEP-1997; 0111166.
 PR 30-JAN-1997; FR-001005.
 PA (CNRS) CNRS CENT NAT RECH SCI.
 PI Boulanger P, Hong SS, Karayan L;
 DR WPI; 98-418459/36.
 PT Targeting adenovirus to cell-surface protein - using bifunctional
 PT ligand that interacts with adenovirus fibre and cell-surface protein
 PS Example 1: Page 34; 43pp; French.
 CC PCR primers V42902-03 were used to amplify part of the genome of
 CC adenovirus serotype 5, and introduce NcoI and KpnI sites for cloning. It
 CC is believed that domain alpha 2 major histocompatibility complex class I
 CC (MHC-I) constitutes the primary receptor for serotype C adenoviruses,
 CC and that human fibronectin module III (sic) functions as a co-receptor
 CC or cofactor. The specification describes a method for targeting
 CC adenovirus to cell-surface proteins other than the natural cellular
 CC receptor of the adenovirus using a bifunctional ligand. The bifunctional
 CC ligand comprises a portion capable of interacting with the adenovirus,
 CC a portion capable of interacting with the cell-surface protein, and
 CC optionally a spacer between the first and second portions. The ligand is
 CC used to target an adenovirus to a host cell of interest, e.g. a tumour
 CC cell, an infected cell, or a particular cell type bearing a specific
 CC surface marker.
 SQ Sequence 36 BP; 6 A; 9 C; 7 G; 14 T;

Query Match 76.0%; Score 15.2; DB 1; Length 36;
 Best Local Similarity 85.0%; Pred. No. 22;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
 DB 5 TTACATTTGACTCTTCCATG 24

RESULT 14
 ID T04199 standard; cDNA to mRNA; 1891 BP.
 AC T04199;
 DT 25-JAN-1996 (first entry)
 DE Potato citrate synthase cDNA.
 KW Citrate synthase; flower formation; tuber storage; ss.
 OS Solanum tuberosum.
 FH Key Location/Qualifiers
 FT cds 73..1485
 FT /*tag= a
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschuetze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR P-PSDB; R82838.

DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Claim 5: Page 53-56; 87pp; English.
 CC To identify a cDNA from potato which codes for citrate synthase, a
 CC cDNA fragment of citrate synthase from Arabidopsis thaliana was
 CC firstly amplified using A.thaliana cDNA and oligos T04202 and T04203
 CC which are complementary to the 5' or 3' end of the coding region of

CC A. thaliana cDNA for citrate synthase. The oligos additionally
 CC introduce BamHI cleavage sites at both ends of the amplified cDNA
 CC fragment. A cDNA library was prep'd. from potato leaves and screened
 CC with A. thaliana citrate synthase cDNA. Positive clones were
 CC purified and sequenced. The nt sequence is given in T04199.
 SQ Sequence 1891 BP; 512 A; 370 C; 425 G; 584 T;

Query Match 74.0%; Score 14.8; DB 1; Length 1891;
 Best Local Similarity 88.9%; Pred. No. 56;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ACATATGAGCCTTCCATG 20
 DB 688 ACATATGAGGATTCATG 705

RESULT 15
 ID T04201 standard; cDNA to mRNA; 1747 BP.
 AC T04201;
 DT 25-JAN-1996 (first entry)
 DE Tobacco citrate synthase cDNA.
 KW Citrate synthase; flower formation; ss.
 OS Nicotiana tabacum
 FH Key Location/Qualifiers
 FT cds 70..1476
 FT /*tag= a
 PN W09524487-A.
 PD 14-SEP-1995.
 PF 07-MAR-1995; E00859.
 PR 09-MAR-1994; DE-408629.
 PR 22-SEP-1994; DE-435366.
 PR 19-OCT-1994; DE-438821.
 PA (AGRE) HOECHST-SCHERING AGREVO GMBH.
 PI La Cognata U, Landschuetze V, Muller-rober B, Landschuetze V;
 PI Mueller-roeber B;
 DR WPI; 95-328278/42.
 DR P-PSDB; R82840.
 PT DNA encoding plant citrate synthase - used to regulate flower formation,
 PT to improve storage of tubers, etc. and to reduce sprouting
 PS Claim 7: Page 60-63; 87pp; English.
 CC To identify a cDNA from tobacco which codes for citrate
 CC synthase, a cDNA bank of leaf tissue from tobacco was prep'd.
 CC plaques of this cDNA bank were screened using a radioactive DNA
 CC probe which comprises Solanum tuberosum citrate synthase cDNA
 CC (T04199). One of the clones was sequenced. The nt. sequence is
 CC given in T04201.
 SQ Sequence 1747 BP; 490 A; 335 C; 400 G; 522 T;

Query Match 74.0%; Score 14.8; DB 1; Length 1747;
 Best Local Similarity 88.9%; Pred. No. 56;
 Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 3 ACATATGAGCCTTCCATG 20
 DB 682 ACATATGAGGATTCATG 699

Search completed: September 18, 1999, 05:27:26
 Job time: 1691 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:46 ; Search time 1405 Seconds
(without alignments)
28.079 Million cell updates/sec

Title: US-09-037-472-8

Perfect score: 20

Sequence: 1 TTACATATGAGCCTTCCATG 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*

1: em_est1:*

2: em_est2:*

3: em_est3:*

4: em_est4:*

5: em_est5:*

6: em_est6:*

7: em_est7:*

8: em_est8:*

9: em_est9:*

10: em_est10:*

11: em_est11:*

12: em_est12:*

13: em_est13:*

14: em_est14:*

15: em_est15:*

16: em_est16:*

17: em_est17:*

18: em_est18:*

19: em_est19:*

20: gb_est1:*

21: gb_est2:*

22: gb_est3:*

23: gb_est4:*

24: gb_est5:*

25: gb_est6:*

26: gb_est7:*

27: gb_est8:*

28: gb_est9:*

29: gb_est10:*

30: gb_est11:*

31: gb_est12:*

32: gb_est13:*

33: gb_est14:*

34: gb_est15:*

35: gb_est16:*

36: gb_est17:*

37: gb_est18:*

38: gb_est19:*

39: gb_est20:*

40: gb_est21:*

41: gb_est22:*

42: gb_est23:*

43: gb_est24:*

44: gb_est25:*

45: gb_est26:*

46: gb_est27:*

47: gb_est28:*

48: gb_est29:*

49: gb_est30:*

50: gb_est31:*

51: gb_est32:*

52: em_est20:*

53: em_est21:*

54: em_est22:*

55: em_est23:*

56: em_est24:*

57: em_est25:*

58: em_est26:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
C 1	17.4	87.0	407	21	R19433
C 2	16.8	84.0	360	23	D55860
C 3	16.8	84.0	476	24	N29325
C 4	16.8	84.0	488	39	AA831249
C 5	16.8	84.0	493	40	AA908884
C 6	16.8	84.0	598	42	AI130847
C 7	16.8	84.0	567	44	AI248140
C 8	16.4	82.0	379	23	D52258
C 9	16.4	82.0	402	23	D53341
C 10	15.8	79.0	447	24	H90734
C 11	15.8	79.0	424	24	H95824
C 12	15.8	79.0	630	30	AA197351
C 13	15.8	79.0	427	31	AA275311
C 14	15.8	79.0	432	31	AA288092
C 15	15.8	79.0	372	32	AA377278
C 16	15.8	79.0	659	37	AA673147
C 17	15.8	79.0	418	38	AA748850
C 18	15.8	79.0	525	38	AA790389
C 19	15.8	79.0	485	39	AA863883
C 20	15.8	79.0	475	43	AI226166
C 21	15.4	77.0	405	23	R93797
C 22	15.4	77.0	474	24	H78857
C 23	15.4	77.0	396	33	AA401166
C 24	15.2	76.0	354	20	M61970
C 25	15.2	76.0	400	21	D46229
C 26	15.2	76.0	490	22	H08527
C 27	15.2	76.0	438	22	H14725
C 28	15.2	76.0	464	22	R51062
C 29	15.2	76.0	275	22	R66347
C 30	15.2	76.0	301	22	R67490
C 31	15.2	76.0	450	24	H73109
C 32	15.2	76.0	447	24	H82720
C 33	15.2	76.0	326	24	N36040
C 34	15.2	76.0	241	25	N46516
C 35	15.2	76.0	408	25	N62510
C 36	15.2	76.0	446	27	AA028135
C 37	15.2	76.0	416	29	AA126072
C 38	15.2	76.0	501	29	AA194540
C 39	15.2	76.0	414	30	AA270555
C 40	15.2	76.0	320	31	AA280522
C 41	15.2	76.0	474	31	AA290446
C 42	15.2	76.0	256	32	AA379224
C 43	15.2	76.0	371	33	AA434013
C 44	15.2	76.0	317	33	AA442950
C 45	15.2	76.0	461	51	AU069310

ALIGNMENTS

RESULT 1

R19433/c 1
LOCUS R19433 407 bp mRNA
DEFINITION y925c03.r1 Soares infant brain INIB Homo sapiens cDNA clone
IMAGE:33575 5', mRNA sequence.
ACCESSION R19433
NID 9773043
VERSION R19433.J GI:773043

EST 14-APR-1995

```

KEYWORDS      EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS        1 (bases 1 to 407)
               Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
               Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
               Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
               Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
               Wilson, R.
TITLE          The WashU-Merck EST Project
JOURNAL        Unpublished (1995)
COMMENT        Contact: Wilson RK
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               Insert Size: 2026
               High quality sequence stops: 228 Source: IMAGE Consortium, LLNL
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               Insert Length: 2026 Std Error: 0.00
               Seq primer: M13Rpi
               High quality sequence stop: 228.
               Location/Qualifiers
               1..407
               /organism="Homo sapiens"
               /db_xref="GDB:405922"
               /db_xref="taxon:9606"
               /clone="IMAGE:33575"
               /clone_lib="Soares infant brain lNIB"
               /sex="female"
               /dev_stage="73 days post natal"
               /lab_host="DH10B (ampicillin resistant)"
               /note="Organ: whole brain; Vector: Lfamid BA; Site: 1: Not
               I; Site 2: Hind III; 1st strand cDNA was primed with a Not
               I - oligo(dT) primer [5'
               AACTGGAGATTCGGCCGCGAGGAATTTTTTTTTTTTTTTT 3'];
               double-stranded cDNA was ligated to Hind III adaptors
               (Pharmacia), digested with Not I and directionally cloned
               into the Not I and Hind III sites of the Lfamid BA vector.
               Library went through one round of normalization. Library
               constructed by Bento Soares and M.Fatima Bonaldo."
BASE COUNT     129 a 58 c 66 g 150 t
ORIGIN
Query Match    87.0%; Score 17.4; DB 21; Length 407;
Best Local Similarity 94.7%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCAT 19
    |||||
Db 338 TTACATATGAGCCTTCAT 320

RESULT 2
LOCUS          D55860
DEFINITION     HUM404C08B Clontech human fetal brain polyA+ mRNA (5535) Homo
               sapiens cDNA clone GEN-404C08 5', mRNA sequence.
ACCESSION      D55860
NID            970260
VERSION        D55860.1 GI:970260
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 360)

```

```

AUTHORS        Fujiwara, T., Hirano, H., Katagiri, T., Kawai, A., Kuga, Y., Nagata, M.,
               Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H.,
               Takaichi, A., Takeda, S., Watanabe, T., Takahashi, E., Hirai, Y.,
               Maekawa, H., Shin, S. and Nakamura, Y.
TITLE          JOURNAL
COMMENT        Fujiwara et al. (1995)
               Unpublished (1995)
               On Sep 21, 1992 this sequence version replaced gi:279304.
               Contact: Tsutomu Fujiwara
               Otsuka GEN Research Institute
               Otsuka Pharmaceutical Co., Ltd
               463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
               Tel: 0886-65-2888
               Fax: 0886-37-1035
               Insert Length: 802 Std Error: 0.00
               High quality sequence stop: 353.
               Location/Qualifiers
               1..360
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="GEN-404C08"
               /clone_lib="Clontech human fetal brain polyA+ mRNA
               (#6535)"
BASE COUNT     96 a 49 c 63 g 152 t
ORIGIN
Query Match    84.0%; Score 16.8; DB 23; Length 360;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCATG 20
    |||||
Db 313 TTACATATGAGCCTTCATG 294

RESULT 3
LOCUS          N29325
DEFINITION     YW84C11.s1 Soares Placenta_8to9weeks_2NbHP8to9w Homo sapiens cDNA
               clone IMAGE:258932 3', mRNA sequence.
ACCESSION      N29325
NID            g1147845
VERSION        N29325.1 GI:1147845
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
               Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 476)
AUTHORS        Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
               Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
               Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
               Trevasakis, E., Waterston, R., Williamson, A., Wohlmann, P. and
               Wilson, R.
TITLE          The WashU-Merck EST Project
JOURNAL        Unpublished (1995)
COMMENT        On May 18, 1995 this sequence version replaced gi:810971.
               Contact: Wilson RK
               Washington University School of Medicine
               4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
               Tel: 314 286 1800
               Fax: 314 286 1810
               Email: est@watson.wustl.edu
               High quality sequence stops: 278
               Source: IMAGE Consortium, LLNL
               This clone is available royalty-free through LLNL; contact the
               IMAGE Consortium (info@image.llnl.gov) for further information.
               Seq primer: ml3 -40 forward
               High quality sequence stop: 278.
               Location/Qualifiers
               1..476

```

```

/organism="Homo sapiens"
/db_xref="GDB:388638"
/db_xref="taxon:9606"
/clone="IMAGE:258932"
/clone_lib="Soares_placenta_8to9weeks_2NbHP8to9W"
/dev_stage="Two placentaes: one from 8 weeks and another
from 9 weeks post conception"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: placenta; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
IGTACCAATCTGAAGTGGAGCGCGCGCATTTTCTTTTCTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library constructed by Bento Soares and
M. Fatima Bonaldo."
BASE COUNT      180 a   92 c   64 g   139 t   1 others
ORIGIN

      Query Match      84.0%; Score 16.8; DB 24; Length 476;
      Best Local Similarity 90.0%; Pred. No. 36;
      Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
    ||||| ||||| ||||| |||||
Db 87 TTACATATGAGACTTTCATG 106

RESULT 4
AA831249 488 bp mRNA EST 21-APR-1998
LOCUS
DEFINITION
oc73e01.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1355352 3'
similar to SW:YN8H_YEAST P53729 HYPOTHETICAL 48.1 KD PROTEIN IN
SEC12-SSK2 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION
AA831249
NID
92904348
VERSION
AA831249.1 GI:2904348
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 488)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 19, 1998 this sequence version replaced gi:2285618.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 794 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 373.
Location/Qualifiers
1..488
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1355352"
/clone_lib="NCI_CGAP_GCB1"
FEATURES
source

```

```

/tissue_type="germinal center B cell"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTGAAGTGGAGCGCGCGCATTTTCTTTTCTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      183 a   94 c   69 g   142 t
ORIGIN

      Query Match      84.0%; Score 16.8; DB 39; Length 488;
      Best Local Similarity 90.0%; Pred. No. 36;
      Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TTACATATGAGCCTTCCATG 20
    ||||| ||||| ||||| |||||
Db 80 TTACATATGAGACTTTCATG 99

RESULT 5
AA908884 493 bp mRNA EST 23-JUN-1998
LOCUS
DEFINITION
om51f02.s1 NCI_CGAP_GC4 Homo sapiens cDNA clone IMAGE:1550715 3'
similar to SW:YN8H_YEAST P53729 HYPOTHETICAL 48.1 KD PROTEIN IN
SEC12-SSK2 INTERGENIC REGION. ;, mRNA sequence.
ACCESSION
AA908884
NID
93048289
VERSION
AA908884.1 GI:3048289
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 493)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL
Unpublished (1997)
COMMENT
On Jan 17, 1998 this sequence version replaced gi:2044779.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 905 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 467.
Location/Qualifiers
1..493
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1550715"
/clone_lib="NCI_CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified

```

polylinker; 1st strand cDNA was prepared from 3 pooled germ cell tumors, and was then primed with a Not I - oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library is normalized. Library was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT 193 a 94 c 67 g 139 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 40; Length 493;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
||||| ||||| ||||| |||||

Db 78 TTACATATGAGACTTTCATG 97

RESULT 6

LOCUS A1130847 598 bp mRNA EST 27-OCT-1998
DEFINITION qc1a10.x1 Soares_fetal_heart_NbHH19W Homo sapiens cDNA clone
IMAGE:1709562 3' similar to SW:YN9H_YEAST P53729 HYPOTHETICAL 48.1
KD PROTEIN IN SEC12-SSK2 INTERGENIC REGION. ;, mRNA sequence.

ACCESSION A1130847
NID q3600863
VERSION A1130847.1 GI:3600863
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

1 (bases 1 to 598)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL

COMMENT

On Jan 9, 1998 this sequence version replaced gi:930548.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 771 Std Error: 0.00

Seq primer: -40ml3 fwd. Et from Amersham

High quality sequence stop: 440.

Location/Qualifiers

FEATURES

source

1..598
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1709562"
/clone_lib="Soares_fetal_heart_NbHH19W"
/sex="unknown"
/dev_stage="19 weeks"

/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a
modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st
strand cDNA was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTGGAGCGCGCATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by
M.Fatima Bonaldo. This library was constructed from the
same fetus as the fetal lung library, Soares fetal lung
NbHL19W."

BASE COUNT 211 a 123 c 88 g 175 t 1 others

ORIGIN

Query Match 84.0%; Score 16.8; DB 42; Length 598;
Best Local Similarity 90.0%; Pred. No. 37;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20
||||| ||||| ||||| |||||

Db 80 TTACATATGAGACTTTCATG 99

RESULT 7

LOCUS A1248140 567 bp mRNA EST 01-DEC-1998
DEFINITION qh75b12.x1 Soares_fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
clone IMAGE:1850495 3' similar to SW:YN9H_YEAST P53729 HYPOTHETICAL
48.1 KD PROTEIN IN SEC12-SSK2 INTERGENIC REGION. ;, mRNA sequence.

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 567)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)

JOURNAL

COMMENT

On Jan 19, 1998 this sequence version replaced gi:2151692.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 772 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 354.

Location/Qualifiers

FEATURES

source

1..567
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="lepl3.3"
/clone="IMAGE:1850495"
/clone_lib="Soares_fetal_liver_spleen_INFLS_S1"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
This is a subtracted version of the original Soares fetal
liver spleen INFLS library. 1st strand cDNA was primed
with a Pac I - oligo(dT) primer [5'
AACTGGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT

ORIGIN

205 a 109 c 82 g 169 t 2 others

Query Match 84.0%; Score 16.8; DB 44; Length 567;
Best Local Similarity 90.0%; Pred. No. 36;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TTACATATGAGCCTTCCATG 20

||||| ||||| ||||| |||||

Db 83 TTACATATGAGACTTTCATG 102


```

RESULT 15
AA377278/c
LOCUS
DEFINITION AA377278 372 bp mRNA EST 21-APR-1997
            EST89818 Small intestine II Homo sapiens cDNA 3' end, mRNA
            sequence.
ACCESSION AA377278
NID 92029596
VERSION AA377278.1 GI:2029596
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 372)
AUTHORS Adams, M.D., Soares, M.B., Kerlavage, A.R., Fields, C. and Venter, J.C.
TITLE Rapid cDNA sequencing (expressed sequence tags) from a
        directionally cloned human infant brain cDNA library
JOURNAL Nature Genet. 4, 373-380 (1993)
MEDLINE 94004965
COMMENT On Sep 12, 1996 this sequence version replaced gi:1288980.
        Other_ESTs: EST89819 THC137556
        Contact: Kerlavage, AR
        Bioinformatics
        The Institute for Genomic Research
        9712 Medical Center Drive, Rockville, MD 20850 USA
        Tel: 3018699056
        Fax: 3018699423
        Email: arkerlav@tigr.org
        For clone availability, additional sequence and expression
        information related to this EST, please check the TIGR Human Gene
        Index (http://www.tigr.org/tdb/hgi/hgi.html)
        Seq primer: M43-21.
FEATURES
        source
            1..372
                /organism="Homo sapiens"
                /db_xref="ATCC (inhost):181679"
                /db_xref="taxon:9606"
                /clone_lib="Small intestine II"
                /dev_stage="adult"
                /note="Organ: small intestine; Vector: pBluescript SK-;
                Site_1: EcoRI; Site_2: XhoI"
                Site_1 68 c 72 g 114 t 3 others
BASE COUNT 115 a 68 c 72 g 114 t 3 others
ORIGIN

Query Match 79.0%; Score 15.8; DB 32; Length 372;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0:

QY 2 TACATATGAGCCTTCATG 20
   ||||| ||||| ||||| |||||
Db 314 TACATTGAGGCTTCATG 296

Search completed: September 18, 1999, 06:47:49
Job time: 3666 sec

```

JOURNAL
COMMENT
University
Unpublished (1999)
On Jul 28, 1997 this sequence version replaced gi:2065396.

Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu

Plate: 605055 row: G column: 07.
Location/Qualifiers

source

/organism="Zea mays"
/cultivar="Ohio43"
/db_xref="taxon:4577"
/clone_lib="605 - Endosperm cDNA library from Schmidt lab"
/tissue_type="pericarp, embryo, and endosperm"
/dev_stage="10 days post-pollination"
/lab_host="DH5(alpha)"
/note="Organ: Kernel; Vector: pAD-GAL4-2'; Site_1: EcoRI;
Site_2: XhoI; Kernel endosperm cDNA library from Schmidt
lab"

BASE COUNT 161 a 180 c 123 g 139 t
ORIGIN

Query Match 73.6%; Score 16.2; DB 50; Length 603;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGTCTACCACTGACTAGG 21
||||| ||||| ||
Db 104 TGTCTACTACCTGACATG 84

RESULT 15

AA390812 466 bp mRNA EST 28-NOV-1998
LOCUS
DEFINITION LD09712.5prime LD Drosophila melanogaster embryo BlueScript
Drosophila melanogaster cDNA clone LD09712 5prime, mRNA sequence.

ACCESSION AA390812

NID 92043958

VERSION AA390812.1 GI:2043958

KEYWORDS EST.

SOURCE fruit fly.

ORGANISM

Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;

Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

1 (bases 1 to 466)

Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Brokstein, P., Lewis, S. and Rubin, G.M.

BDGP/HMI Drosophila EST Project

Unpublished (1997)

On Sep 12, 1996 this sequence version replaced gi:1406966.

REFERENCE

AUTHORS Contact: Harvey, D.

G. M. Rubin-Molecular and Cell Biology

University of California Berkeley

539 USA, Berkeley, CA 94720-3200, USA

Fax: 510 643 9947

Email: http://www.fruitfly.org/EST_est@fruitfly.berkeley.edu

Plate: 97 row: A column: 12

High quality sequence stop: 458.

FEATURES

source

1..466
/organism="Drosophila melanogaster"
/db_xref="BDGP_EST:BDcln008999"
/db_xref="taxon:7227"
/clone="LD09712"
/clone_lib="LD Drosophila melanogaster embryo BlueScript"

/sex="male and female"
/dev_stage="0 to 24 hours mixed stage embryonic"
/lab_host="SOLR"
/note="Organ: embryo; Vector: BlueScript SK; Site_1:
EcoRI; Site_2: XhoI; Constructed using Stratagene ZAP-cDNA
Synthesis kit. Oligo dt-primed and directionally cloned at
EcoRI and XhoI in BlueScript SK(+/-)"
BASE COUNT 139 a 115 c 109 g 103 t
ORIGIN

Query Match 72.7%; Score 16; DB 33; Length 466;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GTTCTACCACTGAAC 17

||||| ||||| |||||

Db 77 GTTCTACCACTGAAC 92

Search completed: September 18, 1999, 06:47:46
Job time: 3663 sec

Query Match 73.6%; Score 16.2; DB 38; Length 452;
 Best Local Similarity 85.7%; Pred. No. 1.9e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGAAGTAGG 21
 ||||| ||||| ||||| |||||
 Db 408 TGTCTACCACTGTGCTATG 388

RESULT 12
 AI374544/c
 LOCUS
 DEFINITION MEST1-D3.POLYTN.Seq ISUM1 Zea mays cDNA clone MEST1-D3 5', mRNA EST 21-JAN-1999
 sequence.
 ACCESSION AI374544
 NID 94174564
 VERSION AI374544.1 GI:4174564
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.

REFERENCE 1 (bases 1 to 909)
 AUTHORS Wen.T.J., Ashlock,D.A. and Schnable,P.S.
 TITLE Expressed Sequence Tags from B73 Maize Seedlings
 JOURNAL Unpublished (1997)
 COMMENT On Jan 17, 1998 this sequence version replaced gi:1900949.

Contact: Schnable, PS
 Schnable laboratory
 Iowa State University
 G405 Agronomy, Ames, IA 50011, USA
 Tel: (515)-294-0975
 Fax: (515)-294-2299
 Email: schnable@iastate.edu
 PCR PRIMERS
 FORWARD: EK1 (5'-CGACGACGACGAAGCCCA-3')
 BACKWARD: HIS1 (5'-GTAGTGGTGGTGGTG-3')
 Plate: MEST1 row: D column: 3
 Seq primer: POLYT-N (5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT(AGC)-3')T7-1.

FEATURES

source
 1..909
 Location/Qualifiers
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /map="22"
 /clone="MEST1-D3"
 /clone_lib="ISUM1"
 /tissue_type="above ground tissues"
 /dev_stage="two-week-old green seedling"
 /lab_host="BL21(DE3)"
 /note="Organ: green seedlings; Vector: pET30a; Site_1: ECORI; Site_2: XhoI; ds-cDNA molecules were generated as follows. First-strand cDNA was prepared from oligo-dT selected mRNA by priming with an XhoI oligo-dT primer. The resulting DNA:RNA hybrid was treated with RNase H and used as a template for DNA PolI-catalyzed second strand synthesis. After the addition of ECORI adaptors, the ds-cDNAs were digested with XhoI and size-selected. The resulting molecules were directionally cloned into the ECORI and XhoI sites of the pET30a plasmid (Novagen)."

BASE COUNT 224 a 241 c 203 g 218 t 23 others
 ORIGIN

Query Match 73.6%; Score 16.2; DB 45; Length 909;
 Best Local Similarity 85.7%; Pred. No. 2.2e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGAAGTAGG 21
 ||||| ||||| ||||| |||||
 Db 128 TGTCTACTACCTGAACAATG 108

RESULT 13
 AI637149/c

LOCUS
 DEFINITION 603001C11.xl 603 - stressed root cDNA library from Wang/Bohnert lab
 Zea mays cDNA, mRNA sequence.
 ACCESSION AI637149
 NID 94688479
 VERSION AI637149.1 GI:4688479
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Zea.
 REFERENCE 1 (bases 1 to 620)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford University
 JOURNAL Unpublished (1999)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3189385.

Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 2227
 Fax: 650 725 8221
 Email: walbot@stanford.edu
 Plate: 603001 row: C column: 11.

FEATURES

source
 1..620
 Location/Qualifiers
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="603 - stressed root cDNA library from Wang/Bohnert lab"
 /tissue_type="seedling"
 /dev_stage="salt stress"
 /lab_host="E. coli XL Gold"
 /note="Organ: root; Vector: pBluescriptII SK(+) XR; Seedling stressed root cDNA library from Wang/Bohnert lab"

BASE COUNT 173 a 186 c 118 g 143 t
 ORIGIN

Query Match 73.6%; Score 16.2; DB 49; Length 620;
 Best Local Similarity 85.7%; Pred. No. 2.1e+02;
 Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGAAGTAGG 21
 ||||| ||||| ||||| |||||
 Db 139 TGTCTACTACCTGAACAATG 119

RESULT 14
 AI677476/c

LOCUS
 DEFINITION 603055G07.xl 605 - Endosperm cDNA library from Schmidt lab Zea mays
 cDNA, mRNA sequence.
 ACCESSION AI677476

NID 94887377
 VERSION AI677476.1 GI:4887377
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.

Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
 Poaceae; Zea.
 REFERENCE 1 (bases 1 to 603)
 AUTHORS Walbot,V.
 TITLE Maize ESTs from various cDNA libraries sequenced at Stanford

```

/db_xref="taxon:6239"
/clone="yk85d9"
/clone_lib="Yuji Kohara unpublished cDNA"
/note="dev stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
BASE COUNT      112 a      68 c      81 g      98 t      1 others
ORIGIN

Query Match      73.6%; Score 16.2; DB 24; Length 360;
Best Local Similarity 85.7%; Pred. No. 1.8e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGAACTAGG 21
      ||||| ||| ||||| |||||
Db 206 TGTCTCCAGCTGACTAGG 186

RESULT 10
AA063903      524 bp      mRNA      EST      03-FEB-1997
LOCUS      m140a10.r1 Stratagene mouse testis (#937308) Mus musculus cDNA
DEFINITION      clone IMAGE:514458 5', similar to gb:M11560 FRUCTOSE-BISPHOSPHATE
ALDOLASE A (HUMAN); gb:Y00516 Mouse mRNA for aldolase A (MOUSE);,
mRNA sequence.
ACCESSION      AA063903
NID      g1557894
VERSION      AA063903.1 GI:1557894
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus.
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS      Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 524)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Scheilenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
On May 5, 1995 this sequence version replaced gi:798078.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:308306
Seq primer: -28ml3 revl ET from Amersham
High quality sequence stop: 59.
Location/Qualifiers
1. 524
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/map="19"
/clone="IMAGE:514458"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTT 3'"
BASE COUNT      128 a      138 c      151 g      107 t

ORIGIN

Query Match      73.6%; Score 16.2; DB 28; Length 524;
Best Local Similarity 85.7%; Pred. No. 2e+02;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GTTCTACCACTGAACTAGGC 22
      ||||| ||||| ||||| ||
Db 192 GTATACCAACCGGAACATGC 212

RESULT 11
AA808387      452 bp      mRNA      EST      12-FEB-1998
LOCUS      oa90d09.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1319537 3',
DEFINITION      similar to TR:Q06265 Q06265 AUTOANTIGEN PM-SCL. [1] ;, mRNA
sequence.
ACCESSION      AA808387
NID      92877793
VERSION      AA808387.1 GI:28777793
KEYWORDS      EST.
SOURCE      human.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 452)
NCI_CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
On Jan 19, 1998 this sequence version replaced gi:2153347.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
Ph.D., Gerald Marti, M.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 217.
Location/Qualifiers
1. 452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1319537"
/clone_lib="NCI_CGAP_GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
/note="vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+ Igd-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCTCAAGTGGGCGGCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      102 a      93 c      77 g      180 t

ORIGIN

```

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbbrp/image/image.html

Seq primer: -40UP from Gibco
High quality sequence stop: 363.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

1. .376
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2306208"
/clone_lib="NCI_CGAP_GC6"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Plasmid DNA from the normalized library
NCI_CGAP_GC4 was prepared, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from a pool of 5,000 clones made
from the same library (cloneIDs 1257096-1258631,
1459064-1470983, and 1475592-1476743). Subtraction by
Bento Soares and M. Fatima Bonaldo."

	BASE COUNT
	ORIGIN

BASE COUNT	78 a	102 c	107 g	89 t
GIN				
Query Match		76.4%	Score 16.8;	DB 49; Length 376;
Best Local Similarity		90.0%;	Pred. No. 93;	
Matches	18;	Conservative	0; Mismatches	2; Indels 0; Gaps 0;

QY 3 TTCTACCACCTGAAC TAGGC 22
||||| ||||| ||||| |||||

Db 213 TTCTCCCACCTGGACTAGGC 232

RESULT 8
B09403/C

LOCUS	R09403	380 bp	EST	05-APR-1995
DEFINITION	Y722809 rl Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:127624 5' mRNA sequence.			
ACCESSION	R09403			
NID	9761326			
VERSION	R09403.1	GI:761326		
KEYWORDS	EST.			
SOURCE	human.			

ORGANISM

REFERENCE
AUTHORS

Eutheria: Primates: Catarrhini; Hominoidea: Homo.
1 (bases 1 to 380)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Huffman, M., Kucab, T., Le, M., Lennon, G., Marr, M.,
Parsons, J., Rifkin, L., Robling, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and
Wilson, R.

TITLE
The WashU-Merck EST Project
JOURNAL
Unpublished (1995)

COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence steps: 282 Source: IMAGE Consortium, LLNL This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 378
Seq primer: M13RP1
Seq Error: 0.00

seq primer: M13KPI
High quality sequence stop: 282.

FEATURES

```

1. 380
source
/organism="Homo sapiens"
/db_xref="GBS:479785"
/db_xref="taxon:9606"
/clone="IMAGE:127624"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Organ: Liver and Spleen; Vector: p77T
with a modified polylinker; Site_1: S
1st strand cDNA was primed with a Pac I - ol
[5'- RACTCGAAGATTAATTAAGAATCTTTTITTTTTT
double-stranded cDNA was ligated to Eco RI a
(Pharmacia), digested with Pac I and cloned
and Eco RI sites of the modified p77T3 vector
went through one round of normalization. Lib
constructed by Bento Soares and M.Fatima Bon
103 a 84 c 91 q 100 t 2 others
BASE COUNT

```

BASE COUNT
ORIGIN

Query Match	73.6%	Score 16.2:	DB 21:	Length 380:
Best Local Similarity	85.7%	Pred. No. 1.9e+03:		
Matches	18:	Conservative	0:	Mismatches 3:
		Indels	0:	Gaps 0:

Qy 1 TGTTCACCACTGAAGTAGG 21
|||||
Db 79 TGTTCACCGCCTGAAGGAGG 59

RESULT 9

LOCUS	D74792	360 bp	EST	14-DEC-1995
DEFINITION	CELK085D9F Yuji Kohara unpublished cDNA <i>Caenorhabditis elegans</i> cDNA clone vk85d9 5', mRNA sequence.			

ACCESSION	D74792
NID	g1120577
VERSION	D74792.1
KEYWORDS	FST
	GI:1120577

KEYWORDS
SOURCE
ORGANISM

Eukaryota: Metazoa: Nematoda: Secernentea: Rhabditia: Rhabditida: Rhabditina: Rhabditidae: Rhabditidae: Peloderinae: Caenorhabditis. 1 (bases 1 to 360)
Kohara, Y., Mitsuiki, H., Nishigaki, A., Motohashi, T., Sugimoto, A. and Tabara, H.

TITLE Toward an expression map of the *C.elegans* genome

JOURNAL Unpublished (1994)
COMMENT On Sep 21, 1992 this sequence version replaced qi:276019.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771

Tel: 0559-75-0771
Fax: 0559-75-6240

Email: vkohara@ddbji.nig.ac.jp

Insert Length: 1084 Std Error: 0.00

High quality sequence stop: 1.

Location/Qualifiers

1. .360

```
/organism="Caenorhabditis elegans"  
/strain="CB1489 him-8(el489)"
```


YPL157W. ; mRNA sequence.
 AA174934
 NID
 G1756082
 VERSION
 AA174934.1 GI:1756082
 KEYWORDS
 EST.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 482)
 REFERENCE
 AUTHORS
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE
 The WashU-HMI Mouse EST Project
 JOURNAL
 Unpublished (1996)
 COMMENT
 On Sep 12, 1996 this sequence version replaced gi:1288871.

Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:378492
 Seq primer: -28M13 rev2 from Amersham
 High quality sequence stop: 455.
 Location/Qualifiers
 1..482

FEATURES

Source
 1..482
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /map="1q"
 /clone="IMAGE:617668"
 /clone_lib="Soares mouse 3NbMS"
 /sex="male"
 /tissue_type="Spleen"
 /dev_stage="4 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was primed with a Not I - oligo(dT) primer [5',
 TGTACCAATCTGAAGTGGAGCGCGCGCTTTTTTTTTTTTTTTTTTT
 3']; double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. RNA
 provided by Dr. Bertrand Jordan. Library went through
 three rounds of normalization, and was constructed by
 Bento Soares and M.Fatima Bonaldo."
 BASE COUNT 160 a 96 c 111 g 115 t
 ORIGIN

Query Match 76.4%; Score 16.8; DB 29; Length 482;
 Best Local Similarity 90.0%; Pred. No. 98;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTTCTACCACTGAATAG 20
 || |||||
 Db 209 TGATCTACCACTGAATAG 228
 RESULT 4
 A1161663/c
 LOCUS
 DEFINITION
 A004P750 Hybrid aspen plasmid library Populus tremula x Populus
 tremuloides cDNA 5', mRNA sequence.
 ACCESSION
 A1161663
 NID
 g3852948

VERSION
KEYWORDS
SOURCE
ORGANISM

A1161663.1 GI:3852948
 EST.
 Populus tremula x Populus tremuloides.
 Populus tremula x Populus tremuloides
 Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
 euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
 eudicots; Rosidae; eucosids I; Malpighiales; Salicaceae; Populus.
 1 (bases 1 to 465)
 REFERENCE
 AUTHORS
 Sterky,F., Regan,S., Karlsson,J., Hertzberg,M., Rohde,A.,
 Holmberg,A., Anini,B., Bhalerao,R., Larsson,M., Villarroel,R., Van
 Montagu,M., Sandberg,G., Olsson,O., Teeri,T., Boerjan,W.,
 Gustafsson,P., Uhlen,M., Sundberg,B. and Lundberg,J.
 TITLE
 Gene discovery in the wood-forming tissues of poplar: Analysis of
 5,692 expressed sequence tags
 JOURNAL
 Proc. Natl. Acad. Sci. U.S.A. 95 (22), 13330-13335 (1998)
 MEDLINE
 99007314
 COMMENT
 On Jan 9, 1998 this sequence version replaced gi:930585.

Contact: Sterky F
 Department of Biotechnology
 Royal Institute of Technology (KTH)
 Teknikringen 34, S-100 44 STOCKHOLM, Sweden
 Tel: +46 8 790 8287
 Fax: +46 8 24 54 52
 Email: fredrik@biochem.kth.se
 PCR Primers
 FORWARD: AAAGGGGATGTGCTGCAAGCG
 BACKWARD: GCTTCGGCTGCTATGTGTGTG
 Seq primer: CGTTGTAACGACGCGCCAG
 High quality sequence stop: 465.
 Location/Qualifiers
 1..465

FEATURES

Source
 1..465
 /organism="Populus tremula x Populus tremuloides"
 /db_xref="taxon:47654"
 /clone_lib="Hybrid aspen plasmid library"
 /tissue_type="Cambial region"
 /dev_stage="1.5 m actively growing tree"
 /lab_host="E.coli"
 /note="Vector: pBluescript SK; Site_1: SalI; Site_2: NotI;
 Cambial region tissues, including developing xylem, the
 meristematic cambial zone and the developing and mature
 phloem, was harvested from 1.5 m actively growing trees.
 cDNA was prepared and cloned into lambda gt2a. DNA was
 isolated and subcloned into pBluescript SK using SalI and
 NotI restriction enzymes."
 BASE COUNT 120 a 101 c 120 g 119 t 5 others
 ORIGIN

Query Match 76.4%; Score 16.8; DB 43; Length 465;
 Best Local Similarity 90.0%; Pred. No. 97;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTTCTACCACTGAATAG 20
 |||||
 Db 150 TGTTCCACCCTGAATCG 131

RESULT 5
LOCUS

A1526675
 DEFINITION
 uJ41e06.y1 Sugano mouse kidney mKia Mus musculus cDNA clone
 IMAGE:1922530 5' similar to WP:T08G11.4 CE13449 ;, mRNA sequence.
 ACCESSION
 A1526675
 NID
 94440810
 VERSION
 A1526675.1 GI:4440810
 KEYWORDS
 EST.
 SOURCE
 house mouse.
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 709)
 REFERENCE
 AUTHORS
 Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,

AA860156.1 GI:2954151
EST.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 443)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1398118.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Possible reversed clone: similarity on wrong strand
Insert Length: 1127 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 76.
Location/Qualifiers
1. .443
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1409081"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from Clontech
Laboratories, Inc., and primed with a Not I - oligo(dT)
primer [5'
TGTTACCAATCTGAAGTGGGAGCGCGCCCAATTTTTTTTTTTT 3'].
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization to Cot5, and was
constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 113 a 90 c 104 g 136 t
ORIGIN
Query Match 78.2%; Score 17.2; DB 39; Length 443;
Best Local Similarity 86.4%; Pred. No. 60;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGTCTACCACTGAAGTGGC 22
||||||| || |||||
Db 266 TGTCTACGCCAGACTAGGC 245

RESULT 2
LOCUS AI558929 238 bp mRNA EST 24-MAR-1999
DEFINITION fb78b02.y1 zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
WP:T04C9.6 CE04881 BAND 4.1-LIKE DOMAIN ; mRNA sequence.
ACCESSION AI558929
NID q4509167
VERSION AI558929.1 GI:4509167
SOURCE EST.
ORGANISM zebrafish.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;

Cyprinoidae; Cyprinidae; Rasborinae; Danio.
1 (bases 1 to 238)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M.,
Eddy, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T.,
Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y.,
Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R.,
Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R.,
Waterson, R. and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3188593.

Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
CDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center clone distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 106.
Location/Qualifiers
1. .238
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue.type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XL1-blue MRF"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGACATGTTCTAGATCGGAGCGCGCGCTTTTITTTT3'];
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."

BASE COUNT 73 a 58 c 48 g 59 t
ORIGIN
Query Match 78.2%; Score 17.2; DB 48; Length 238;
Best Local Similarity 86.4%; Pred. No. 53;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 1 TGTCTACCACTGAAGTGGC 22
||||||| |||||
Db 183 TGTCTACCACTGAAGTGGC 204

RESULT 3
LOCUS AA174934 mRNA EST 16-FEB-1997
DEFINITION ms78c03.r1 Soares mouse 3NDMS Mus musculus cDNA clone IMAGE:517668
5' similar to TR:E246895 E246895 CHROMOSOME XVI READING FRAME ORF

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:43 ; Search time 1405 Seconds
(without alignments)
30.887 Million cell updates/sec

Title: US-09-037-472-7

Perfect score: 22

Sequence: 1 TGTTCACCACTGAAGTAGGC 22

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	17.2	78.2	443	39	AA860156	AA860156 ak47b09.s
C 2	17.2	78.2	238	48	AI558929	AI558929 fb78b02.y
C 3	16.8	76.4	482	29	AA174934	AA174934 ms78c03.r
C 4	16.8	76.4	465	43	AI161663	AI161663 A004P750
C 5	16.8	76.4	709	47	AI526675	AI526675 u44e06.y
C 6	16.8	76.4	484	49	AI526755	AI526755 ms78c03.y
C 7	16.8	76.4	376	49	AI652206	AI652206 wb20c01.x
C 8	16.2	73.6	380	21	R09403	R09403 yf22e09.r1
C 9	16.2	73.6	360	24	D74792	D74792 CELK085D9F
C 10	16.2	73.6	524	28	AA063903	AA063903 ml40a10.r
C 11	16.2	73.6	452	38	AA808387	AA808387 oa90d09.s
C 12	16.2	73.6	909	45	AI374544	AI374544 MEST1-D3.
C 13	16.2	73.6	620	49	AI637149	AI637149 603001C11
C 14	16.2	73.6	603	50	AI677476	AI677476 605055G07
C 15	16.2	72.7	466	33	AA390812	AA390812 LD09712.5
C 16	16.2	72.7	465	38	AA441586	AA441586 LD16566.5
C 17	16.2	72.7	692	40	AA950076	AA950076 LD30041.5
C 18	15.8	71.8	419	34	AA497993	AA497993 vi69g04.r
C 19	15.8	71.8	488	40	AA971375	AA971375 op80d11.s
C 20	15.8	71.8	384	43	AI210120	AI210120 g9f06a1.r
C 21	15.8	71.8	490	43	AU033738	AU033738 AU033738
C 22	15.8	71.8	573	43	AU033743	AU033743 AU033743
C 23	15.8	71.8	689	45	AU000589	AU000589 AU000589
C 24	15.8	71.8	637	45	AU004377	AU004377 AU004377
C 25	15.8	71.8	603	47	AI520078	AI520078 LD40035.5
C 26	15.6	70.9	403	25	N50071	N50071 yz11a06.sl
C 27	15.6	70.9	405	25	N88530	N88530 K3598F_Huma
C 28	15.6	70.9	338	25	W11649	W11649 ma92b05.r1
C 29	15.6	70.9	960	25	W13530	W13530 ma85d01.r1
C 30	15.6	70.9	1077	26	W20613	W20613 mb90g09.r1
C 31	15.6	70.9	448	26	W70002	W70002 zds1b11.r1
C 32	15.6	70.9	187	26	W97061	W97061 mg08b04.r1
C 33	15.6	70.9	199	27	AA004050	AA004050 mg81b02.r
C 34	15.6	70.9	604	27	W91371	W91371 mf94g02.r1
C 35	15.6	70.9	625	27	W98817	W98817 mf94e05.r1
C 36	15.6	70.9	358	28	AA065112	AA065112 zf72d05.s
C 37	15.6	70.9	670	28	AA080216	AA080216 mj99b06.f
C 38	15.6	70.9	297	28	AA082350	AA082350 ze88h08.r
C 39	15.6	70.9	471	28	AA098401	AA098401 mo12d08.r
C 40	15.6	70.9	608	28	AA107288	AA107288 mp05a04.r
C 41	15.6	70.9	481	28	AA109871	AA109871 mm02e07.r
C 42	15.6	70.9	648	29	AA167738	AA167738 zq40e03.s
C 43	15.6	70.9	188	29	AA183193	AA183193 mt80e01.r
C 44	15.6	70.9	411	30	AA237610	AA237610 mx14g12.r
C 45	15.6	70.9	714	53	HSM000559	AI036251 Homo sapi

ALIGNMENTS

RESULT 1
AA860156/c
LOCUS AA860156 443 bp mRNA
DEFINITION ak47b09.s1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1409081
3' similar to SW:RSU1_HUMAN Q15404 RAS SUPPRESSOR PROTEIN 1 ; , mRNA
sequence.
ACCESSION AA860156
NID g2954151

SQ Sequence 335 BP; 94 A; 68 C; 94 G; 79 T;

Query Match 70.9%; Score 15.6; DB 1; Length 335;
Best Local Similarity 81.8%; Pred. No. 51;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGACTAGGC 22
||| ||||| ||||| ||| ||| |||
Db 206 TGCTCTACCACTGAGCTATGC 185

RESULT 15

V52331
ID V52331 standard; DNA; 6846 BP.
AC V52331;
DT 23-OCT-1998 (first entry)
DE Streptococcus pneumoniae genome fragment SEQ ID NO:198.
KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KW computer readable medium; vaccine; pharmaceutical composition; ds.
OS Streptococcus pneumoniae.
PN WO/818931-A2.
PD 07-MAY-1998.
PF 30-OCT-1997; U19588.
PR 31-OCT-1996; US-029960.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI Kunsch CA, Rosen CA;
DR WPI: 98-27225/24.
PT Computer-readable medium with recorded Streptococcus pneumoniae
PT polynucleotide sequences - useful in diagnostic kits and assays, and
PT pharmaceutical compositions and vaccines for Streptococcus
PT pneumoniae
PS Claim 1; Page 1169-1173; 1409pp; English.
CC The present invention describes a computer readable medium which has
CC the nucleotide sequences SEQ ID NO:1 to 391 (V52134 to V52524) recorded
CC on it, or a representative fragment or a sequence at least 95% identical
CC to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in SEQ ID NO:1
CC to 391 (V52134 to V52524) are genomic fragments from Streptococcus
CC pneumoniae. The present invention also describes an isolated nucleic acid
CC molecule encoding a homologue of any of the fragments of the S.pneumoniae
CC genome (SEQ ID NO:1 to 391) where the nucleic acid molecule is produced
CC by a process comprising: (a) screening a genomic DNA library using as a
CC probe a target sequence defined by any of the sequences in SEQ ID NO:1
CC to 391, identifying members of the library which contain sequences
CC that hybridize to the target sequence and isolating the nucleic acid
CC molecules from the members; or (b) isolating mRNA, DNA or cDNA produced
CC from an organism, amplifying nucleic acid molecules whose nucleotide
CC sequence is homologous to amplification primers derived from the
CC fragment of the S. pneumoniae genome to prime the amplification and
CC isolating the amplified sequences. The computer readable medium can be
CC used in a computer-based system for identifying fragments of the
CC S. pneumoniae genome of commercial importance, or expression modulating
CC fragments of the S. pneumoniae genome. Products from the present
CC invention can be used in diagnosis kits and assays, and pharmaceutical
CC compositions and vaccines for S. pneumoniae.
SQ Sequence 6846 BP; 2142 A; 1588 C; 1139 G; 1972 T;

Query Match 70.9%; Score 15.6; DB 1; Length 6846;
Best Local Similarity 81.8%; Pred. No. 74;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGACTAGGC 22
||| ||||| ||||| ||| ||| |||
Db 6569 TGCTCTACCACTGAGCTAAGC 6590

Search completed: September 18, 1999, 05:27:24
Job time: 1689 sec

DT 20-JAN-1993 (first entry)
DE Encodes exons XV to XVIII of human hepatocyte growth factor.
KW HGF; enhance growth; preparing transgenic animals; hepatic disease;
KW clinical diagnostic reagent; drug.
OS Homo sapiens.
FH Key Location/Qualifiers
FT 101..241
FT /*tag= a
FT /note= "exon XV"
FT 773..879
FT /*tag= b
FT /note= "exon XVI"
FT 991..1136
FT /*tag= c
FT /note= "exon XVII"
FT 3764..3937
FT /*tag= d
FT /note= "exon XVIII"
FT poly_a_signal 7492..7497
FT /*tag= e
PN J04183394-A.
PD 30-JUN-1992.
PF 19-NOV-1990; 314548.
PR 19-NOV-1990; JP-314548.
PA (NAKA/) NAKAMURA T.
PA (TOYM) TOYOBO KK.
DR WPI; 92-265591/32.
DR P-PSDB; R25692.
PT Recombinant human hepatocyte growth factor and DNA encoding it -
PT useful for diagnosis and treatment of hepatic disease and
PT transgenic animal prepn.
PS Disclosure; Page 22; 28pp; Japanese.
CC This sequence contains exons XV to XVIII of human hepatocyte growth
CC factor. See also R25676-92, Q26713-27.
SQ Sequence 7753 BP; 2567 A; 1177 C; 1243 G; 2766 T;

Query Match 71.8%; Score 15.8; DB 1; Length 7753;
Best Local Similarity 89.5%; Pred. No. 60;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGTCTTACCACCTGAACCTA 19
|||
DB 6269 TGTCTTACCACCTGAACCTA 6251

RESULT 12
Q14102/c
ID Q14102 standard; DNA: 603 BP.
AC Q14102;
DT 10-JAN-1992 (first entry)
DE N.gonorrhoeae strain NCTC 8375 16S to 23S rRNA gene spacer region.
KW rRNA gene; ribosomal RNA; probe; ss.
OS Neisseria gonorrhoeae NCTC 8375.
PN EP-452596-A.
PD 23-OCT-1991.
PF 18-APR-1990; 401054.
PR 18-APR-1990; EP-401054.
PA (INNO-) INNOGENETICS NV SA.
PI Rossau R, Van Heuverswijn H;
DR WPI: 91-311940/43.
PT Hybridisation probes for detecting non-viral microorganisms -
PT derived from spacer region between 16S and 23S rRNA genes, for
PT detecting e.g. BORDETELLA PERTUSSIS, NEISSERIA GONORRHOEA
PS Disclosure; Fig 3; 41pp; English.
CC This sequence is the non-coding strand of the 16S-23S rRNA gene
CC spacer region; the 5'-end is proximal to the 16S rRNA gene and the 3'
CC -end proximal to the 23S rRNA gene. The sequence is very similar
CC to the corresponding region in N.gonorrhoeae ITM 4367. A set of
CC probes was designed based on this sequence which was specific for
CC N.gonorrhoeae. A kit is provided for detection of N.gonorrhoeae
CC using the probes.
SQ Sequence 603 BP; 203 A; 104 C; 145 G; 151 T;

Query Match 70.9%; Score 15.6; DB 1; Length 603;
Best Local Similarity 81.8%; Pred. No. 55;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGTCTTACCACCTGAACCTAGGC 22
|||
DB 206 TGTCTTACCACCTGAACCTAGGC 185

RESULT 13
Q14106/c
ID Q14106 standard; DNA: 664 BP.
AC Q14106;
DT 10-JAN-1992 (first entry)
DE N.meningitidis NCTC 10025 16S to 23S rRNA gene spacer region.
KW rRNA gene; ribosomal RNA; probe; ss.
OS Neisseria meningitidis NCTC 10025.
PN EP-452596-A.
PD 23-OCT-1991.
PF 18-APR-1990; 401054.
PR 18-APR-1990; EP-401054.
PA (INNO-) INNOGENETICS NV SA.
PI Rossau R, Van Heuverswijn H;
DR WPI: 91-311940/43.
PT Hybridisation probes for detecting non-viral microorganisms -
PT derived from spacer region between 16S and 23S rRNA genes, for
PT detecting e.g. BORDETELLA PERTUSSIS, NEISSERIA GONORRHOEA
PS Disclosure; Fig 3; 41pp; English.
CC This sequence is the non-coding strand of the 16S-23S rRNA gene
CC spacer region; the 5'-end is proximal to the 16S rRNA gene and the 3'
CC -end proximal to the 23S rRNA gene. A set of probes was designed
CC based on this sequence which was specific for N.meningitidis. A kit
CC is provided for detection of this species using the probes.
SQ Sequence 664 BP; 235 A; 123 C; 154 G; 152 T;

Query Match 70.9%; Score 15.6; DB 1; Length 664;
Best Local Similarity 81.8%; Pred. No. 55;
Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGTCTTACCACCTGAACCTAGGC 22
|||
DB 208 TGTCTTACCACCTGAACCTAGGC 187

RESULT 14
Q14103/c
ID Q14103 standard; DNA: 335 BP.
AC Q14103;
DT 10-JAN-1992 (first entry)
DE N.gonorrhoeae strain ITM 4367 16S to 23S rRNA gene spacer region.
KW rRNA gene; ribosomal RNA; probe; ss.
OS Neisseria gonorrhoeae ITM 4367.
PN EP-452596-A.
PD 23-OCT-1991.
PF 18-APR-1990; 401054.
PR 18-APR-1990; EP-401054.
PA (INNO-) INNOGENETICS NV SA.
PI Rossau R, Van Heuverswijn H;
DR WPI: 91-311940/43.
PT Hybridisation probes for detecting non-viral microorganisms -
PT derived from spacer region between 16S and 23S rRNA genes, for
PT detecting e.g. BORDETELLA PERTUSSIS, NEISSERIA GONORRHOEA
PS Disclosure; Fig 1; 41pp; English.
CC This sequence is the non-coding strand of the 16S-23S rRNA gene
CC spacer region; the 5'-end is proximal to the 16S rRNA gene and the 3'
CC -end proximal to the 23S rRNA gene. The sequence differs in just one
CC position from the corresponding region in N.gonorrhoeae NCTC 8375,
CC i.e. at position 143. A set of probes were designed based on this
CC sequence which were specific for N.gonorrhoeae. A kit is provided
CC for detection of N.gonorrhoeae using the probes.

SQ Sequence 4190 BP; 780 A; 1406 C; 1306 G; 698 T;

Query Match 73.6%; Score 16.2; DB 1; Length 4190;
Best Local Similarity 85.7%; Pred. No. 35;
Matches 18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 GTTCTACCACTGAAGTGGC 22
||||||| ||||| |||||

Db 3522 GTTCTACCTCTGCACGAGC 3502

RESULT 8

V77989

ID V77989 standard; DNA: 648 BP.

AC V77989.

DT 16-MAR-1999 (first entry)

DE Staphylococcus aureus contig SEQ ID #3678.

KW Computer readable medium; vaccine; S. aureus infection; immunodetection;

KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;

KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

FH Key Location/Qualifiers

FT 121. .180

FT misc_feature

FI /*tag= a

FI /note= "these bases represent a line of missing text in

the sequence listing in the specification. They

are included to maintain the nucleotide numbering

given in the specification for this DNA sequence"

FN EP-786519-A2.

PD 30-JUL-1997.

PF 07-JAN-1997; 100117

PR 05-JAN-1996; US-009861

PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,

PI Rosen CA;

DR WPI; 97-374922/35.

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -

stored on computer readable medium and used in the production of

anti-S. aureus vaccines

PS Claim 1; Page 2639-2640; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences

of the invention. The DNA sequences are recorded on a computer readable

medium, preferably selected from a floppy or hard disk, random access

memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

the S. aureus DNA sequences allows putative functions to be assigned so

that protein-encoding or regulatory regions of commercial, therapeutic or

industrial importance can be obtained. Specifically, sequences which are

likely to encode antigens have been identified and these polypeptides can

be used in a vaccine composition against S. aureus infection. The

polypeptides can also be used in a kit for the immunodetection of

S. aureus in a sample. S. aureus is implicated in numerous human diseases,

including cellulitis, eyelid infections, food poisoning, osteomyelitis,

skin and surgical wound infections, scalded skin syndrome, toxic shock

syndrome, etc. Organisms transformed with the DNA sequences can be used

for recombinant production of the polypeptides. The new DNA sequences

(and their fragments) are useful as primers or probes for isolating

homologues of any of the S. aureus DNA sequences contained on the

computer readable medium.

SQ Sequence 648 BP; 161 A; 150 C; 144 G; 131 T;

Query Match

Best Local Similarity, 73.6%; Score 16.2; DB 1; Length 648;

Matches 18; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGAAGTGGC 22

||||||| ||||| |||||

Db 10 TGNCTACCAATTGAGCTAGC 31

RESULT 9

Q26727/c

ID Q26727 standard; DNA: 7753 BP.

AC Q26727;

X20248_03/c

Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotid
WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

Fragment Name

WP X20248_00

WP X20248_01

WP X20248_02

WP X20248_03

WP X20248_04

WP X20248_05

WP X20248_06

WP X20248_07

WP X20248_08

WP X20248_09

WP X20248_10

WP X20248_11

WP X20248_12

WP X20248_13

WP X20248_14

WP X20248_15

WP X20248_16

WP X20248_17

WP X20248_18

WP X20248_19

WP X20248_20

WP X20248_21

WP X20248_22

WP X20248_23

WP X20248_24

WP X20248_25

WP X20248_26

WP X20248_27

WP X20248_28

WP X20248_29

WP X20248_30

WP X20248_31

WP X20248_32

WP X20248_33

WP X20248_34

WP X20248_35

WP X20248_36

WP X20248_37

WP X20248_38

WP X20248_39

WP X20248_40

WP X20248_41

WP X20248_42

WP X20248_43

WP X20248_44

WP X20248_45

WP X20248_46

WP X20248_47

WP X20248_48

WP X20248_49

WP X20248_50

WP X20248_51

WP X20248_52

WP X20248_53

WP X20248_54

WP X20248_55

WP X20248_56

WP X20248_57

WP X20248_58

WP X20248_59

WP X20248_60

WP X20248_61

WP X20248_62

WP X20248_63

WP X20248_64

WP X20248_65

WP X20248_66

WP X20248_67

WP X20248_68

WP X20248_69

WP X20248_70

WP X20248_71

WP X20248_72

WP X20248_73

WP X20248_74

WP X20248_75

WP X20248_76

WP X20248_77

WP X20248_78

WP X20248_79

WP X20248_80

WP X20248_81

WP X20248_82

WP X20248_83

WP X20248_84

WP X20248_85

WP X20248_86

WP X20248_87

WP X20248_88

WP X20248_89

WP X20248_90

WP X20248_91

WP X20248_92

WP X20248_93

WP X20248_94

WP X20248_95

WP X20248_96

WP X20248_97

WP X20248_98

WP X20248_99

WP X20248_100

WP X20248_101

WP X20248_102

WP X20248_103

WP X20248_104

WP X20248_105

WP X20248_106

WP X20248_107

WP X20248_108

WP X20248_109

WP X20248_110

WP X20248_111

WP X20248_112

WP X20248_113

WP X20248_114

WP X20248_115

WP X20248_116

WP X20248_117

WP X20248_118

WP X20248_119

WP X20248_120

WP X20248_121

WP X20248_122

WP X20248_123

WP X20248_124

WP X20248_125

WP X20248_126

WP X20248_127

WP X20248_128

WP X20248_129

WP X20248_130

WP X20248_131

WP X20248_132

WP X20248_133

WP X20248_134

WP X20248_135

WP X20248_136

WP X20248_137

WP X20248_138

WP X20248_139

WP X20248_140

WP X20248_141

WP X20248_142

WP X20248_143

WP X20248_144

WP X20248_145

WP X20248_146

WP X20248_147

WP X20248_148

WP X20248_149

WP X20248_150

WP X20248_151

WP X20248_152

WP X20248_153

WP X20248_154

WP X20248_155

WP X20248_156

WP X20248_157

WP X20248_158

WP X20248_159

WP X20248_160

WP X20248_161

WP X20248_162

WP X20248_163

WP X20248_164

WP X20248_165

WP X20248_166

WP X20248_167

WP X20248_168

WP X20248_169

WP X20248_170

WP X20248_171

WP X20248_172

WP X20248_173

WP X20248_174

WP X20248_175

WP X20248_176

WP X20248_177

WP X20248_178

WP X20248_179

WP X20248_180

WP X20248_181

WP X20248_182

WP X20248_183

WP X20248_184

WP X20248_185

WP X20248_186

WP X20248_187

WP X20248_188

WP X20248_189

WP X20248_190

WP X20248_191

WP X20248_192

WP X20248_193

WP X20248_194

WP X20248_195

WP X20248_196

WP X20248_197

WP X20248_198

WP X20248_199

WP X20248_200

WP X20248_201

WP X20248_202

WP X20248_203

WP X202

AC X13096;
 DT 19-MAR-1999 (first entry)
 DE Enterococcus faecalis genome contig SEQ ID NO:159.
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 OS Enterococcus faecalis.
 PN W09850555-A2.
 PD 12-NOV-1998.
 PF 04-MAY-1998; U008985.
 PR 14-NOV-1997; US-066009.
 PR 06-MAY-1997; US-044031.
 PR 16-MAY-1997; US-046655.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Dillon PJ, Kunsch CA.
 DR WPI: 99-045171/04.
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 PS Claim 1: Page 898-902; 2084pp; English.
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC X12938 to X13919 represent these nucleotide sequences which are primary
 CC nucleotide sequences, also known as contigs. The computer-based system
 CC can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 SQ Sequence 8160 BP; 2731 A; 1463 C; 1654 G; 2292 T;

Query Match 76.4%; Score 16.8; DB 1; Length 8160;
 Best Local Similarity 90.0%; Pred. No. 19;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 TGTCTACCACTGAAGTAG 20
 DB 3366 TGTCTAGCTCTGAAGTAG 3385

RESULT 6
 ID V75647/c
 DT V75647;
 AC V75647;
 DE 16-MAR-1999 (first entry)
 KW Staphylococcus aureus contig SEQ ID #1336.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 PN EP-786519-A2.
 PD 30-JUL-1997.
 PF 07-JAN-1997; 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI: 97-374922/35.
 PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1: Page 1924; 3271pp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 633 BP; 252 A; 105 C; 90 G; 184 T;

Query Match 74.5%; Score 16.4; DB 1; Length 633;
 Best Local Similarity 94.4%; Pred. No. 22;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 TGTCTACCACTGAAGTAG 18
 DB 85 TGTCTACCACTGAAGTAG 68

RESULT 7
 ID V34854/c
 DT V34854;
 AC V34854;
 DE 28-SEP-1998 (first entry)
 KW Human retinal degeneration B2 polynucleotide (hrdgB2).
 KW Non-receptor tyrosine kinase binding protein; screening; brain; heart;
 KW thymus; leukocytes; human fetal brain; PIR domain; Pyk2 binding domain;
 KW synaptic vesicles; neurotransmitter signalling; myasthenia gravis;
 KW stroke; neuroblastoma; thrombocytopaenia; Alzheimer's; Huntington's;
 KW Parkinson's; depression; schizophrenia; pain epilepsy; ss.
 OS Homo sapiens.
 FH Key
 FT CDS
 FT 174..3908
 FT /*tag= a
 FT /product= "hrdgB2 protein"
 FT W09816639-Al.
 PN 23-APR-1998.
 PD 26-SEP-1997; U17374.
 PR 11-OCT-1996; US-027337.
 PA (SUG-) SUGEN INC.
 PA (UYNY-) UNIV NEW YORK MEDICAL CENT.
 PI Lev S, Plovman GD, Schlessinger J;
 DR WPI: 98-251286/22.
 DR P-PSDB; W59358.
 PT New nucleic acid encoding human retinal degradation polypeptide -
 PT and related probes, vectors, transformed cells, proteins and
 PT antibodies, used to diagnose neurological disease and to identify
 PT specific modulators with possible therapeutic activity
 PS Disclosure; page 36-37; 59pp; English.
 CC The present sequence represents the human retinal degeneration B2 cDNA
 CC (hrdgB2), the product of which is a non-receptor tyrosine kinase binding
 CC protein. The isolation of this cDNA involved the amplification of an EST
 CC fragment (T12574) from human fetal brain cDNA. This product was
 CC subcloned, sequenced and then used as a probe to screen a human fetal
 CC brain cDNA library. The clone obtained was used as a probe to rescreen
 CC the same library, from this seven clones were found, subcloned and
 CC sequenced. These clones were found to be of the gene hrhgB2. hrhgB2
 CC protein has been found to be expressed in the brain, heart, thymus and
 CC peripheral blood leukocytes. It contains a PIT domain, and a Pyk2 binding
 CC domain and is thus concerned with recycling synaptic vesicles and
 CC regulating neurotransmitter signalling respectively. This protein is
 CC seen to be involved in signalling transduction pathways and therefore
 CC would be useful in diagnosis, treatment, and prevention of the following
 CC diseases: myasthenia gravis; neuroblastoma; thrombocytopaenia; stroke;
 CC Alzheimer's; Huntington's; Parkinson's; depression; schizophrenia;
 CC pain epilepsy.

PT Predicting increased risk of sight-threatening diabetic retinopathy
PT - comprises identifying genetic polymorphism pattern for genes
PT IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
PT symptoms occur
PS Claim 2; Page 33; 41pp; English.
CC Interleukin-1-alpha (IL-1-alpha) primers 1 and 2 (V32390) were used to
CC amplify the IL-1-alpha gene region to identify single base variation
CC polymorphism of C/T at base 889. The invention claims to provide a
CC method for predicting the risk of sight threatening diabetic retinopathy.
CC The method involves isolating DNA from a patient and determining the DNA
CC polymorphism pattern of the genes that code for interleukin-1-alpha,
CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
CC identified is then compared with controls of known DNA polymorphism
CC patterns thereby identifying patients carrying a genetic polymorphism
CC associated with increased risk of sight threatening diabetic retinopathy.
CC The method may be able to identify diabetic patients at risk before the
CC clinically detectable disorders occur. Polymorphism pattern
CC determination of IL genes involved PCR reactions using primers V32389-
CC V32398. The method is also claimed to be useful in conjunction with
CC identification of other genes associated with sight threatening diabetic
CC retinopathy in genomic DNA and therefore, in identifying diabetic
CC patients expressing multiple risk patterns.
SQ Sequence 27 BP; 7 A; 8 C; 5 G; 7 T;

Query Match 100.0%; Score 22; DB 1; Length 27;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTCTTACCACCTGAAGTAGGC 22
|||||
DB 6 TGTCTTACCACCTGAAGTAGGC 27

RESULT 3

ID X02988 standard; DNA: 5701 BP.
AC X02988;
DT 22-JUN-1999 (first entry)
DE Human IL-1ra BAC contiguous DNA sequence 33.
KW TANGO-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
PN WO9906426-A1.
PD 11-FEB-1999.
PE 03-AUG-1998; U16102.
PR 02-JUL-1998; US-091650.
PR 04-AUG-1997; US-054546.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Pan Y;
PT NP; 99-153692/13.
DR New isolated nucleic acid encoding the new human cytokine TANGO-77 -
DR used to inhibit inflammation and to screen for specific modulators
PS Example 5; Figure 3; 226pp; English.
CC X02956-X03048 and X23301-X22304 are overlapping BAC genomic sequences
CC containing alternatively spliced forms of human IL-1ra. Such fragments
CC are used in the method of the invention which describes the isolation of
CC a novel human TANGO-77 encoding nucleic acid and protein. TANGO-77 is a
CC member of the cytokine superfamily that is expected to inhibit
CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC also bind to a new receptor so could regulate other cellular processes
CC associated with acute or chronic inflammation, e.g. asthma, chronic
CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent
CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 5701 BP; 1729 A; 1238 C; 1079 G; 1654 T;

Query Match 100.0%; Score 22; DB 1; Length 5701;
X13096

Best Local Similarity 100.0%; Pred. No. 0.049;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTCTTACCACCTGAAGTAGGC 22
|||||
DB 431 TGTCTTACCACCTGAAGTAGGC 452

RESULT 4

ID X16611 standard; DNA: 21 BP.
AC X16611;
DT 29-APR-1999 (first entry)
DE Interleukin 1 (44112332) haplotype PCR primer #5.
KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;
KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;
KW Graves disease; systemic lupus erythematosus; lichen sclerosis;
KW ulcerative colitis; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9854359-A1.
PD 03-DEC-1998.
PE 21-MAY-1998; G01481.
PR 29-MAY-1997; GB-011040.
PA (CAMP/) CAMP N J.
PA (COXA/) COX A.
PA (DGIO/) DE GIOVINE F S.
PA (DUFF/) DUFF G.
PI Camp NJ, Cox A, De Giovine FS, Duff G;
PD WPI; 99-080814/07.

PT New method of determining a patient's susceptibility to inflammatory
PT disorders - by detecting the presence of an IL-1 (44112332)
PT haplotype, useful in designing treatment strategies that modulate
PT the activity of proteins produced by the IL-1 gene cluster
PS Claim 3; Page 33; 49pp; English.
CC A method has been developed for determining a patient's susceptibility
CC to an inflammatory disorder. The method comprises the detection of an
CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the
CC patient, where its presence indicates susceptibility to an inflammatory
CC disorder. X16607 to X16631 represent PCR primer used in the method for
CC detecting the IL-1 (44112332) haplotype. The method provides kits for
CC the early prediction of a patient's susceptibility to inflammatory
CC disorders, including coronary artery disease, osteoporosis, nephropathy
CC in diabetes mellitus, alopecia areata, Graves disease, systemic lupus
CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of
CC alleles of the haplotype can be applied to particular inflammatory
CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,
CC psoriasis, and insulin dependent diabetes. The identification of a
CC disease-associated haplotype enables determination of which alleles are
CC causative, and this information is useful in designing treatment
CC strategies, including gene therapy and treatment using various agents
CC that modulate the activity of proteins produced by the IL-1 gene cluster.
CC Some alleles from the IL-1 gene cluster are associated with particular
CC inflammatory diseases, and insufficient IL-1 production appears to act
CC centrally in the pathology of these diseases. Therefore, the use of IL-1
CC gene clusters is useful in determining genetic susceptibility to
CC inflammatory diseases, including those with a multifactorial etiology
CC with a polygenic component.
SQ Sequence 21 BP; 5 A; 6 C; 4 G; 6 T;

Query Match 95.5%; Score 21; DB 1; Length 21;
Best Local Similarity 100.0%; Pred. No. 0.075;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTCTTACCACCTGAAGTAGG 21
|||||
DB 1 TGTCTTACCACCTGAAGTAGG 21

RESULT 5

X13096
ID X13096 standard; DNA: 8160 BP.

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 05:27:23 ; Search time 213.04 Seconds
(without alignments)
25.837 Million cell updates/sec

Title: US-09-037-472-7

Perfect score: 22

Sequence: 1 TGTCTACCACTGAAGTAGGC 22

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	22	100.0	22	1 T70316	Primer for detecti
2	22	100.0	27	1 V32389	Interleukin-1-alph
3	22	100.0	5701	1 X02988	Human IL-1ra BAC c
4	21	95.5	21	1 X16611	Interleukin 1 (441
5	16.8	76.4	8160	1 X13096	Enterococcus faeca
6	16.4	74.5	633	1 V75647	Staphylococcus aur
7	16.2	73.6	4190	1 V34854	Human retinal dege
8	16.2	73.6	648	1 V77989	Staphylococcus aur
9	16.2	73.6	110000	1 X20248_03	Continuation (4 of
10	16.2	73.6	73	1 X30293	Borrelia burgdorfe
11	15.8	71.8	7753	1 Q26727	Encodes exons XV t
12	15.6	70.9	603	1 Q14102	N.gonorrhoeae str
13	15.6	70.9	664	1 Q14106	N.meningitidis NCT
14	15.6	70.9	335	1 Q14103	N.gonorrhoeae str
15	15.6	70.9	6846	1 V52331	Streptococcus pneu
16	15.6	70.9	19718	1 V52232	Streptococcus pneu
17	15.6	70.9	59	1 V79270	Staphylococcus aur
18	15.6	70.9	400	1 V78290	Staphylococcus aur
19	15.6	70.9	361	1 V78258	Staphylococcus aur
20	15.6	70.9	400	1 V78053	Staphylococcus aur
21	15.6	70.9	400	1 V78055	Staphylococcus aur
22	15.6	70.9	401	1 V78042	Staphylococcus aur
23	15.6	70.9	553	1 V78001	Staphylococcus aur
24	15.6	70.9	400	1 V77971	Staphylococcus aur
25	15.6	70.9	400	1 V77942	Staphylococcus aur
26	15.6	70.9	475	1 V77932	Staphylococcus aur
27	15.6	70.9	1171	1 V77850	Staphylococcus aur
28	15.6	70.9	400	1 V77853	Staphylococcus aur
29	15.6	70.9	6591	1 V77425	Staphylococcus aur
30	15.6	70.9	237	1 V78877	Staphylococcus aur
31	15.6	70.9	239	1 V78855	Staphylococcus aur
32	15.6	70.9	239	1 V78854	Staphylococcus aur
33	15.6	70.9	327	1 V78683	Staphylococcus aur
34	15.6	70.9	411	1 V78684	Staphylococcus aur
35	15.6	70.9	339	1 V78574	Staphylococcus aur
36	15.6	70.9	400	1 V78545	Staphylococcus aur
37	15.6	70.9	381	1 V78334	Staphylococcus aur
38	15.6	70.9	400	1 V78340	Staphylococcus aur
39	15.6	70.9	9797	1 X13487	Enterococcus faeca
40	15.6	70.9	1421	1 X13384	Enterococcus faeca
41	15.6	70.9	22960	1 X13282	Enterococcus faeca
42	15.6	70.9	240	1 X13970	H. pylori GHPD 125
43	15.6	70.9	6422	1 X20576	Polynucleotide seq

44 15.2 69.1 1992 1 Q10895 Encodes Xenopus Bo
45 15.2 69.1 3565 1 Q84051 Sequence encoding

ALIGNMENTS

```
RESULT 1
T70316
ID T70316 standard; DNA; 22 BP.
AC T70316;
DT 19-OCT-1997 (first entry)
DE Primer for detecting genetic predisposition to periodontal disease.
KW Periodontal disease; gingivitis; parodontitis; polymorphism;
KW interleukin-1 alpha; IL-1A; interleukin-1 beta; IL-1B; primer; PCR;
KW polymerase chain reaction; ss.
OS Synthetic.
PN WO9706180-A1.
PD 20-FEB-1997.
PF 02-AUG-1996; U12455.
PR 03-AUG-1995; US-510696.
PA (KORN/) KORNMAN K S.
PI (MEDI-) MEDICAL SCI SYSTEMS INC.
PI Duff GW, Kornman KS;
DR WPI: 97-154207/14.
PT Identification of patient's genetic polymorphism pattern - allows
PT prediction of increased periodontal disease severity
PS Claim 3; Page 25; 33pp; English.
CC PCR primers (T70316 and T13884-88) can be used in a new method
CC for identifying a genetic predisposition to periodontal disease by
CC detecting the presence of DNA polymorphisms in the gene sequences
CC for interleukin-1 alpha (IL-1A) and interleukin-1 beta (IL-1B).
CC Alleles associated with severe disease were identified as IL-1A
CC allele 2 together with IL-1B (TaqI) allele 2. The primers are used
CC to amplify DNA from a blood or tissue sample and the products are
CC subjected to restriction digestion to determine the polymorphism
CC pattern of the patient. A single base variation (C/T) polymorphism
CC at IL-1A base -889 can be identified using primers (T70316 and
CC T13884) corresponding to bases -967 to -945 and -888 to -869,
CC respectively. An NcoI site is created if C is available at -889,
CC but not if T is present.
SQ Sequence 22 BP; 5 A; 7 C; 4 G; 6 T;

Query Match 100.0%; Score 22; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.024;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 TGTCTACCACTGAAGTAGGC 22
Db 1 TGTCTACCACTGAAGTAGGC 22

RESULT 2
V32389
ID V32389 standard; DNA; 27 BP.
AC V32389;
DT 11-SEP-1998 (first entry)
DE Interleukin-1-alpha primer 1 (-967/-945).
KW IL-1-alpha; genetic polymorphism; PCR; primer; amplification;
KW sight threatening diabetic retinopathy; interleukin-1-alpha;
KW interleukin-1-beta; interleukin-1RN; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9815653-A1.
PD 16-APR-1998.
PF 09-OCT-1997; G02790.
PR 10-OCT-1996; GB-021129.
PA (DUFF/) DUFF G.
PA (RENN/) RENNIE I.
PA (RICH/) RICHARDSON R.
PI Duff G, Rennie I, Richardson R;
DR WPI: 98-240835/21.
```

```

gene      7825.. .8073
/ gene="psaC"
CDS       7825.. .8073
/ gene="psaC"
/codon_start=1
/product="PSI, Fe-S polypeptide SU VII, 9 kDa"
/protein_id="CAA91614.1"
/db_xref="PID:e211807"
/db_xref="PID:g1185131"
/db_xref="GI:1185131"
/db_xref="SWISS-PROT:P49477"
/translation="MSHTVKIYDTICGTCOCVPACPTDULEMPWDGCKSGQTASSPR
VEDCVGCKCEACPTDFLSRVYLGRLRLETRSLGLAY"
complement(8139.. .9077)
/ gene="ycf5"
CDS       complement(8139.. .9077)
/ gene="ycf5"
/codon_start=1
/product="ORF312"
/protein_id="CAA91615.1"
/db_xref="PID:e211808"
/db_xref="PID:g1185132"
/db_xref="GI:1185132"
/db_xref="SWISS-PROT:P49523"
/translation="MDWDLIQNLSNLVEGILLFAMTIYWISLSFFKWTKNLSQVGKI
SAIVANILUFFILGSRWIVAGYFPLSNLYESLLFTLTLLTIYLYVEFKTKSLVGAI
LIPVALLINGFANLTLSVDMQKSSPLPALQSNWLMHVSMMMLSYGTLIMGSLLCIL
FLVISKYQIDILDESSLPLYNIMLDYEAKEKLFSPSDEVSELGKLLQLSDNWSY
RIIGLGFPLTIGIAGGYVANEANGSYWSDPKETWALTITWIVFATYILHSRITKGWE
GKKTALGLGFGFVINICILGVNFGKGLHSGWLS"
9287.. .9427
/ gene="ycf32"
/ note="A-T-rich"
CDS       /codon_start=1
/ product="ORF46"

```

```

Query Match      76.4%; Score 16.8; DB 7; Length 119704;
Best Local Similarity 90.0%; Pred. No. 1.3e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      2 GTTCTACCACTGAACTAGG 21
        |||||
Db 92708 GTTCTACCACTGAACTAG 92727

```

Search completed: September 18, 1999, 15:49:22
Job time: 1442 sec

protein L27; 50S ribosomal protein L29; 50S ribosomal protein L3; 50S ribosomal protein L31; 50S ribosomal protein L32; 50S ribosomal protein L33; 50S ribosomal protein L34; 50S ribosomal protein L35; 50S ribosomal protein L36; 50S ribosomal protein L4; 50S ribosomal protein L5; 50S ribosomal protein L6; 5S ribosomal RNA; 5S RNA gene; acp gene; acyl carrier protein; apocytochrome f; ATP synthase CF1; ATP synthase CFO; atpA gene; atpB gene; atpD gene; atpE gene; atpF gene; atpG gene; atpH gene; atpI gene; cfxQ gene; chaperonin; chlI gene; ctp protease; clpC gene; complete genome; CP43 chlorophyll apoprotein; CP47 chlorophyll apoprotein; cytochrome b559; cytochrome b6; cytochrome b6/f complex; cytochrome c550; D1 reaction-center protein; D2 reaction-center protein; dna-replication helicase; dnaB gene; dnaK gene; elongation factor tu; Fe-S polypeptide; ferredoxin; ferredoxin-binding protein II; groEL gene; orf 179; ORF 312; ORF 644; ORF113; ORF128; ORF148; ORF181; ORF204; ORF251; ORF263; ORF266; ORF267; ORF282; ORF290; ORF296; ORF319; ORF344; ORF355; ORF355 gene; ORF36; ORF380; ORF382; ORF41; ORF42; ORF44; ORF45; ORF46; ORF486; ORF497; ORF61; ORF64; ORF73; ORF74; ORF99; p700 apoprotein A1; p700 apoprotein A2; petA gene; petB gene; petD gene; petF gene; petG gene; petK gene; phosphoprotein; plastocyanin-binding; preprotein-translocase; protein I; protein J; protein K; protein L; protein M; protein N; protein T; protein W; protein X; psaA gene; psaB gene; psac gene; psad gene; psae gene; psaf gene; psal gene; psaj gene; psal gene; psam gene; psba gene; psbb gene; psbc gene; psbd gene; psbe gene; psbf gene; psbh gene; psbi gene; psbj gene; psbk gene; psbl gene; psbn gene; psbr gene; psbv gene; psbw gene; psbx gene; PSI; PSII; rbcL gene; rbcS gene; RNA polymerase; RNA polymerase b'-chain; RNA polymerase b'-chain; RNA polymerase b-chain; rpl1 gene; rpl11 gene; rpl12 gene; rpl13 gene; rpl14 gene; rpl16 gene; rpl18 gene; rpl19 gene; rpl2 gene; rpl20 gene; rpl21 gene; rpl22 gene; rpl23 gene; rpl24 gene; rpl27 gene; rpl29 gene; rpl3 gene; rpl31 gene; rpl32 gene; rpl32' gene; rpl33 gene; rpl34 gene; rpl35 gene; rpl36 gene; rpl4 gene; rpl5 gene; rpl6 gene; rpoA gene; rpoB gene; rpoC1 gene; rpoC2 gene; rps10 gene; rps11 gene; rps12 gene; rps13 gene; rps14 gene; rps16 gene; rps17 gene; rps18 gene; rps19 gene; rps2 gene; rps21 gene; rps3 gene; rps4 gene; rps5 gene; rps6 gene; rps7 gene; rps8 gene; rps9 gene; rubisco large subunit; rubisco small subunit; secA gene; secY gene; thiG gene; trnA gene; trnA-Gln; trnA-Ala; trnA-Arg; trnA-Asn; trnA-Asp; trnA-Cys; trnA-Met; trnA-Glu; trnA-Gly; trnA-His; trnA-Ile; trnA-Leu; trnA-Lys; trnA-Met; trnA-Phe; trnA-Pro; trnA-Ser; trnA-Thr; trnA-Tyr; trnA-Val; trnC gene; trnD gene; trnE gene; trnF gene; trnM gene; trnG gene; trnH gene; trnI gene; trnK gene; trnL gene; trnM gene; trnN gene; trnP gene; trnQ gene; trnR gene; trnS gene; trnT gene; trnV gene; trnW gene; trnY gene; trsE gene; tufA gene; ycf12 gene; ycf16 gene; ycf24 gene; ycf25 gene; ycf3 gene; ycf30 gene; ycf31 gene; ycf32 gene; ycf33 gene; ycf35 gene; ycf39 gene; ycf4 gene; ycf40 gene; ycf41 gene; ycf42 gene; ycf43 gene; ycf44 gene; ycf45 gene; ycf46 gene; ycf47 gene; ycf5 gene; ycf6 gene; ycf7 gene; ycf9 gene; Odontella sinensis

SOURCE
ORGANISM Chloroplast Odontella sinensis
Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae; Biddulphiophycidae; Eupodiscaceae; Eupodiscaceae; Odontella.
REFERENCE
1 (bases 1 to 119704)
AUTHORS Kowallik,K.V., Stoebe,B., Schaffran,I., Kroth-Pancic,P. and Freier,U.
TITLE The Chloroplast Genome of a chlorophyll a+c- containing Alga, Odontella sinensis
JOURNAL Plant Mol. Biol. Rep. 13, 336-342 (1995)
REFERENCE 2 (bases 1 to 119704)
AUTHORS Kowallik,K.V.
TITLE Direct Submission
JOURNAL Submitted (10-NOV-1995) K.V. Kowallik, Heinrich-Heine Universitaet Duesseldorf, Universitaetsstr. 1 Geb.26 13/02/46, D- 40225, Duesseldorf, FRG
FEATURES
source Location/Qualifiers
1. .119704
/organism="Odontella sinensis"
/chloroplast
/db_xref="taxon:2839"
misc_feature 1. .7725

/note="Ira"
267. .340
/gene="trnP"
/product="trNA-Pro (ugg)"
267. .54092
/gene="trnP"
478. .1545
/gene="ORF355"
478. .1545
/gene="ORF355"
/codon_start=1
/protein_id="CAA91611.1"
/db_xref="PID:e217972"
/db_xref="PID:g1185128"
/db_xref="GI:1185128"
/translation="MGMSLQLLGNLRKLVATLSWLFENIAKVCGYPTKILGMPIRT
DIVPGYGPPIHOTRVPPFPANPTTFLEAVGNIPQPTTEKFVYESPODGYNYFPHY
RNVPFLPWLKSKQLHFDLDIDYVGLTINTLFESMLVYFFFAELRIMLSYFISIN
PYTRPWVYLISLTDWIYDILFHLGSKRWLLGFPPLLLIHAALGNLIDSLNHLVFT
MPFLPSEGEPEGLDGTARKVLLFRYLPALWTSEPIPRDRREFWYTERPEIYOFMKK
NYGLHIDFLPDEILKQIYQFKHDQIDPNSLTKNVEQFKVLGAQLISDSTIDFYQFSN
CFVHKQEILFTFFSDYMDRLI"
2210. .3694
/gene="16S rRNA"
/product="16S ribosomal RNA"
2210. .52149
/gene="16S rRNA"
3761. .3835
/gene="trnI"
/product="trNA-Ile (gau)"
3761. .65452
/gene="trnI"
3838. .3910
/gene="trnA"
/product="trNA-Ala (ugc)"
3838. .50521
/gene="trnA"
3976. .5866
/gene="23S rRNA"
/product="23S ribosomal RNA"
3976. .50383
/gene="23S rRNA"
6938. .7063
/gene="5S rRNA"
/product="5S ribosomal RNA"
6938. .47421
/gene="5S rRNA"
7318. .47041
/gene="ycf32"
7318. .7428
/gene="ycf32"
/codon_start=1
/product="ORF36"
/protein_id="CAA91612.1"
/db_xref="PID:e217944"
/db_xref="PID:g1185129"
/db_xref="GI:1185129"
/translation="MDSRLLVIAAPLLVAASWALFNIGRLAQIQRLISA"
7563. .7739
/gene="rpl32"
7563. .7739
/gene="rpl32"
/codon_start=1
/product="50S ribosomal protein L32"
/protein_id="CAA91613.1"
/db_xref="PID:g11806"
/db_xref="PID:g1185130"
/db_xref="GI:1185130"
/db_xref="SWISS-PROT:P49564"
/translation="MAVPKKRTSKAKKNARKSVWKKKADKAAKKSLSLAKSVLOGKTT
SFVSLYIDELFSI"

tRNA
gene
gene
CDS

gene
CDS

gene
CDS

carbon dioxide concentrating mechanism protein CcmK;
 carboxyl-terminal protease; cell division FtsZ protein; cell
 division inhibitor; cell division protein FtsH; chloroplast
 import-associated channel IAP75; chorismate synthase; cysteine
 deaminase; cytoplasmic membrane protein for maltose uptake; cytosine
 6-phosphate dehydrogenase; glutaredoxin 3; glutathione synthetase;
 glycoprotein 64; high light inducible protein; histidinol-phosphate
 aminotransferase; hydrogenase expression/formaton protein HspE;
 lysyl-tRNA synthetase; naphthoate synthase; nitrate reductase;
 nitrate transport 45kD protein; nitrate transport protein NrtB;
 nitrate-chain-release factor 3; phosphoribosyl aminidazole
 succinocarboxamide synthetase; poly(3-hydroxyalkanoate) synthase;
 polypeptide deformylase; protein kinase C inhibitor; protein kinase
 PknA; protochlorophyllide reductase 57 kD subunit; pyridine
 nucleotide transhydrogenase alpha subunit; pyruvate dehydrogenase
 E1 beta subunit; rare lipoprotein A; regulatory components of
 sensory transduction system; secreted protein MPB70; secretory
 protein SecE; succinate dehydrogenase flavoprotein subunit;
 tRNA-Leu; tRNA-Trp; transcription antitermination subunit;
 transposase; urease accessory protein E.
 Synechocystis sp. (strain:PC66803) DNA.
 Synechocystis sp.
 Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 1 (bases 1 to 135551)
 Tabata, S.
 Direct Submission
 Submitted (28-JUN-1996) to the DDBJ/EMBL/GenBank databases. Satoshi
 Tabata. Kazusa DNA Research Institute, Laboratory of Gene Structure
 2; 1523-3, Ianauchino, Kisarazu, Chiba 292, Japan
 (E-mail:tabata@kazusa.or.jp, Tel: +81-438-52-3933,
 Fax: +81-438-52-3934)
 2 (sites)
 Kaneko, T., Sato, S., Kotani, H., Tanaka, A., Asami, E., Nakamura, Y.,
 Miyajima, N., Hikosawa, M., Sugliura, M., Sasamoto, S., Kimura, T.,
 Hosouchi, T., Tabata, S., Muraki, A., Nakazaki, N., Naruo, K.,
 Okumura, S., Shimpo, S., Takeuchi, C., Wada, T., Watanabe, A.,
 Yamada, M., Yasuda, M. and Tabata, S.
 Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions
 DNA Res. 3 (3), 109-136 (1996)
 97061201
 Potential protein coding regions were assigned on the basis of
 similarity search of the ORFs and GeneMark analysis.
 Location/Qualifiers
 1. .135551
 /organism="Synechocystis sp."
 /strain="PC6803"
 /db_xref="taxon:1143"
 complement(27. 281)
 /note="ORF_ID:s13712"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="BAAL7414.1"
 /db_xref="PID:d1018147"
 /db_xref="PID:g1652493"
 /db_xref="GI:1652493"
 /translation="MGETKLTADQRMQGTGTHCDKVILSYIELHRRRYELEVEVAFSWFP
 HVPPEKTLTAAWEKELILKWRSPFNKECWRRWGQPFGLS"
 complement(634. .1722)
 /gene="aroc"
 complement(634. .1722)
 /gene="aroc"
 /note="ORF_ID:s111747"
 /codon_start=1
 /transl_table=11
 /product="chorismate synthase"
 /protein_id="BAAL7415.1"
 /db_xref="PID:d1018148"
 /db_xref="PID:g1652494"

/db_xref="GI:1652494"
 /translation="MGNTFOSLPRITTFGSHGGGVVIIDGCPRLRLEISPEETQVDL
 DRRPQOSKITTPRKADQCEILSGVFEKTLGTPIAILVRNKDARSQDYENMAVKYR
 PSHADATYEAKYIRWQGGSSARETIGRVAAGAIAKKILAQFNGEIVAVVYKIQ
 DTEATVDSNTVITLQEVSNIVRCPDECAKMERIDQVLROKDSIGVVVECAIRNAP
 KTGEPVDFDKLEADAKAMSLPATKGFEGSGFAGTLLTGSOHNDEYILDEAGWRT
 RTNRSGVGGISNGSPIINRIAFKPTATIGQEQKTVSNIGEEITLAAKRHRDPCVLP
 RAYPMVMAALVICDHLRFAQCKTL"
 complement(1929. .2315)
 /gene="rpl12"
 complement(1929. .2315)
 /gene="rpl12"
 /note="ORF_ID:s111746"
 /codon_start=1
 /transl_table=11
 /product="50S ribosomal protein L12"
 /protein_id="BAAL7416.1"
 /db_xref="PID:d1018149"
 /db_xref="PID:g1652495"
 /db_xref="GI:1652495"
 /translation="MSAATDQILEOLKSLLEASELVLKQIEEAFGVSAAPVGGMMV
 AAAAAAPAEAEKTEFDVILVEVPADKKIAVLKVVRTITGLGLKEAKELVESTPKAI
 KEATGKDDAEAIKKQIEEAGKAAVR"
 complement(2403. .2924)
 /gene="rpl10"
 complement(2403. .2924)
 /gene="rpl10"
 /note="ORF_ID:s111745"
 /codon_start=1
 /transl_table=11
 /product="50S ribosomal protein L10"
 /protein_id="BAAL7417.1"
 /db_xref="PID:d1018150"
 /db_xref="PID:g1652496"
 /db_xref="GI:1652496"
 /translation="MGRFRENKATVSDQELFQDAQMTVIIDYOGLTVAEITDLNR
 LRPLQGTCKIAKNTLVRRALAQEAWSPNEEELTGTAILVLKEDLGGAIAKTKFQK
 DTKTELRGVDEGKSLQADVEAIGDLPSQOLMGQIAGGINALATKIALGKEVPA
 SVARGLOAHVDKE"
 complement(3171. .3887)
 /gene="rpl1"
 complement(3171. .3887)
 /gene="rpl1"
 /note="ORF_ID:s111744"
 /codon_start=1
 /transl_table=11
 /product="50S ribosomal protein L1"
 /protein_id="BAAL7418.1"
 /db_xref="PID:d1018151"
 /db_xref="PID:g1652497"
 /db_xref="GI:1652497"
 /translation="MTKKLSKRMQAAIAKVDSDSKLYSPLEAMELLKETATAKFDTA
 EAHIRLIDPKYSQDQIRTTVSLPKGTGTVRVAVILARGEKVKAEATDAGADIAGSELI
 EEIQGMDDFDVLIATPDMPKIAIRLQGLPGLMPSPKGGTVDADLAAYNEFKAG
 KLEFRADRTGIVHVMFGKASFSADLLANLKAQETIDRNRSFGARGFRWRTVTVSSS
 MGSIPVDINALRDLKFEDN"
 complement(3974. .4399)
 /gene="rpl11"
 complement(3974. .4399)
 /gene="rpl11"
 /note="ORF_ID:s111743"
 /codon_start=1
 /transl_table=11
 /product="50S ribosomal protein L11"
 /protein_id="BAAL7419.1"
 /db_xref="PID:d1018152"
 /db_xref="PID:g1652498"
 /db_xref="GI:1652498"
 /translation="MAKKRVALLKALPAGKANPAPPVGPALGOHGNIMAFCKEYNA
 KTADKPGMIIPVEISVFEDRSFTFLKTPPASVILIRKAAGVEKSGSEPNKKNVASITR
 EQLREIAOTKLPDLNANDIDAAMNIEGTARNMGITVNS"
 complement(4501. .5118)
 /gene="nugS"


```

VERSION      X16767.1  GI:44265
KEYWORDS     transfer RNA; transfer RNA-Trp.
SOURCE       Mycoplasma capricolium.
ORGANISM     Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
              capricolium group.
REFERENCE    1 (bases 1 to 76)
AUTHORS      Andachi,Y., Yamao,F., Muto,A. and Osawa,S.
TITLE        Codon recognition patterns as deduced from sequences of the
              complete set of transfer RNA species in Mycoplasma capricolium.
              Resemblance to mitochondria
JOURNAL      J. Mol. Biol. 209 (1), 37-54 (1989)
MEDLINE      90040718
REFERENCE    2 (bases 1 to 76)
AUTHORS      Yamao,F., Iwagami,S., Azumi,Y., Muto,A., Osawa,S., Fujita,N. and
              Ishihama,A.
TITLE        Evolutionary dynamics of tryptophan tRNAs in Mycoplasma capricolium
JOURNAL      Mol. Gen. Genet. 212 (2), 364-369 (1988)
MEDLINE      88302126
COMMENT      [2] see x07691 for genomic sequence
              [2] Data kindly reviewed (27-JUL-88) by YAMAO F.
              See acc# X16741-X16769 for complete set of mycoplasma transfer
              RNA's.

FEATURES     Location/Qualifiers
             source      1..76
                       /organism="Mycoplasma capricolium"
                       /strain="ATCC 27343 (Kid.)"
                       /db_xref="taxon:2095"
             trna         1..76
                       /note="tRNA-Trp"
             modified_base 20
                       /note="dihydrouridine"
                       /mod_base=d
             misc_feature  34..36
                       /note="anticodon (UCA)"
             modified_base  34
                       /note="5-carboxymethylaminomethyl-2'-O-methyluridine"
                       /mod_base=OTHER
             modified_base  37
                       /note="N6-methyladenosine"
                       /mod_base=m6a
             modified_base  46
                       /note="7-methylguanosine"
                       /mod_base=m7g
             modified_base  55
                       /note="pseudouridine"
                       /mod_base=p
             BASE COUNT    18 a 18 c 21 g 19 t
             ORIGIN
             1 TGTCTACCACTGAAGTACGC 22
             26 TGTCTACTACTGAAGTATGC 5

             Query Match      78.2%; Score 17.2; DB 1; Length 76;
             Best Local Similarity 86.4%; Pred. No. 1.4e+02;
             Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGTCTACCACTGAAGTACGC 22
    ||||| ||||| ||||| ||||| ||
Db 26 TGTCTACTACTGAAGTATGC 5

RESULT 9
MYCTGW/c  MYCTGW 994 bp DNA BCT 12-AUG-1994
LOCUS     Mycoplasma capricolium Trp-tRNA genes.
DEFINITION
ACCESSION K02974
NID       g150214
VERSION   K02974.1 GI:150214
KEYWORDS  transfer RNA; transfer RNA-Trp.
SOURCE    M. capricolium (American type culture collection 27343) DNA, clone
              pMCH964.
ORGANISM  Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
              capricolium group.

```

```

REFERENCE    1 (bases 1 to 994)
AUTHORS      Yamao,F., Muto,A., Kawauchi,Y., Iwami,M., Iwagami,S., Azumi,Y. and
              Osawa,S.
TITLE        UGA is read as tryptophan in Mycoplasma capricolium
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 82, 2306-2309 (1985)
MEDLINE     85190486
COMMENT      Draft entry and reprint for [1] kindly provided by F.Yamao,
              20-SEP-1985.
              Even though 'tga' is a nonsense or termination (opal) codon
              throughout prokaryotes and eukaryotes, it codes for tryptophan in
              M.capricolium, as is the case in mitochondria. The sequence of the
              S3 and L16 ribosomal protein genes from this wall-less bacterium
              include 'tga' codons in their reading frames in positions
              corresponding to tryptophan in E.coli S3 and L16 genes. A Trp-tRNA
              with the opal suppressing anticodon 'tca' is found in M.capricolium.
              A promoter region is located at positions 371-399 and an area of
              dyad symmetry, which could be a transcription-termination structure
              is found at positions 614-640.

FEATURES     Location/Qualifiers
             source      1..994
                       /organism="Mycoplasma capricolium"
                       /db_xref="taxon:2095"
                       /db_xref="taxon:2095"
                       /note="codon recognized: TGA; Trp-tRNA"
                       /product="tRNA-Trp"
             trna         529..603
                       /note="Trp-tRNA"
             BASE COUNT    345 a 125 c 145 g 379 t
             ORIGIN
             153 bp upstream of AluI site.

             Query Match      78.2%; Score 17.2; DB 1; Length 994;
             Best Local Similarity 86.4%; Pred. No. 1.2e+02;
             Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 TGTCTACCACTGAAGTACGC 22
    ||||| ||||| ||||| ||||| ||
Db 441 TGTCTACTACTGAAGTATGC 420

RESULT 10
MYCTRGCT/c MYCTRGCT 300 bp DNA BCT 02-FEB-1999
LOCUS     Mycoplasma capricolium tRNA gene cluster encoding two Trp-tRNAs.
DEFINITION
ACCESSION D00551
NID       g216805
VERSION   D00551.1 GI:216805
KEYWORDS  tRNA-Trp.
SOURCE    Mycoplasma capricolium (strain:Kid; ATCC 27343) DNA.
ORGANISM  Mycoplasma capricolium
              Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
              capricolium group.
              1 (bases 1 to 300)
              2 (bases 1 to 300)
              Yamao,F., Muto,A., Kawauchi,Y., Iwami,M., Iwagami,S., Azumi,Y. and
              Osawa,S.
              UGA is read as tryptophan in Mycoplasma capricolium
              Proc. Natl. Acad. Sci. U.S.A. 82 (8), 2306-2309 (1985)
              85190486
              2 (bases 1 to 300)
              Muto,A., Andachi,Y., Yuzawa,H., Yamao,F. and Osawa,S.
              The organization and evolution of transfer RNA genes in Mycoplasma
              capricolium
              Nucleic Acids Res. 18 (17), 5037-5043 (1990)
              90384798
              These data kindly submitted in computer readable form by: Akira
              Muto
              Department of Biology, School of Science, Nagoya University
              Furo-cho, Chikusa-ku
              Nagoya 464-01
              Japan
              Phone: 052-781-5111 x6642
              Fax: 052-783-0719.
              Location/Qualifiers

```

```

modified_base      /mod_base-m7g
54                /mod_base-t
modified_base      55
BASE COUNT        15 a 21 c 23 g 17 t
ORIGIN            5' end of mature tRNA.

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 1; Length 76;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGGAATAGGC 22
Db 26 TGTCTACCAATGGAATAGGC 5

RESULT 6
ECOTRWSUP/c
LOCUS      ECOTRWSUP      76 bp      tRNA      BCT      20-MAY-1994
DEFINITION E.coli Trp-tRNA suppressor mutant.
ACCESSION  M25066
NID        g174466
VERSION    M25066.1 GI:174466
KEYWORDS   transfer RNA-Trp suppressor.
SOURCE     E.coli tRNA.
ORGANISM   Escherichia coli
            Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
            Escherichia.
REFERENCE  1 (bases 1 to 76)
AUTHORS   Eisenberg,S.P., Yarus,M. and Soll,L.
TITLE     The effect of an Escherichia coli regulatory mutation on transfer
            RNA structure
JOURNAL   J. Mol. Biol. 135, 111-126 (1979)
MEDLINE   80117855
COMMENT    In the trpX mutation the base at position 37 ('a') is not modified.
            In Trp-tRNA base 24 'a' is changed to 'g'.
FEATURES   Location/Qualifiers
            source
            1..76
            /organism="Escherichia coli"
            /db_xref="taxon:562"
            tRNA
            1..76
            /note="codon recognized: TGG; Phe-tRNA"
            /product="tRNA-Phe"
modified_base     16
modified_base     17
modified_base     20
modified_base     32
modified_base     37
modified_base     46
modified_base     /mod_base-ms216a
modified_base     /note="gm7"
modified_base     /mod_base-OTHER
modified_base     /note="psi"
BASE COUNT      15 a 21 c 23 g 17 t
ORIGIN          5' end of mature tRNA.

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 1; Length 76;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGGAATAGGC 22
Db 26 TGTCTACCAATGGAATAGGC 5

modified_base     54
modified_base     55
BASE COUNT      15 a 21 c 23 g 17 t
ORIGIN          5' end of mature tRNA.

Query Match
Best Local Similarity 78.2%; Score 17.2; DB 1; Length 281;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGGAATAGGC 22
Db 74 TGTCTACCTACTGGAATATGC 53

RESULT 8
MCTRW2/c
LOCUS      MCTRW2      76 bp      mRNA      BCT      31-MAR-1992
DEFINITION Mycoplasma capricolum transfer RNA-Trp (UCA).
ACCESSION  X16767 X07953
NID        g44265

```

```

RESULT 7
MCTRNW12/c
LOCUS      MCTRNW12      281 bp      DNA      BCT      31-MAR-1992
DEFINITION Mycoplasma capricolum genes for tRNA-Trp (UCA) and tRNA-Trp (CCA).
ACCESSION  X07691
NID        q44254
VERSION    X07691.1 GI:44254
KEYWORDS   transfer RNA; transfer RNA-Trp.
SOURCE     Mycoplasma capricolum.
ORGANISM   Mycoplasma capricolum;
            Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
            capricolum group.
REFERENCE  1 (bases 1 to 281)
AUTHORS   Yamao,F., Iwagami,S., Azumi,Y., Muto,A., Osawa,S., Fujita,N. and
            Ishihama,A.
TITLE     Evolutionary dynamics of tryptophan tRNAs in Mycoplasma capricolum
JOURNAL   Mol. Gen. Genet. 212 (2), 364-369 (1988)
MEDLINE   88302126
COMMENT    see x07953 and x07954 for tRNA sequences
            Data kindly reviewed (27-JUL-1988) by YAMAO F.
FEATURES   Location/Qualifiers
            source
            1..281
            /organism="Mycoplasma capricolum"
            /db_xref="taxon:2095"
            /clone="pMCH964"
            4..9
            /note="put.-35 region"
            27..32
            /note="put.-10 region"
            39..278
            /note="transcript"
            39..158
            /note="transcript"
            49..124
            /note="transfer RNA-Trp (UCA)"
            82..84
            /note="transfer RNA-Trp (UCA) anticodon"
            135..155
            /note="pot.-stem-loop structure; terminator-like structure"
            135..143
            /note="inverted repeat A"
            147..155
            /note="inverted repeat A"
            162..236
            /note="transfer RNA-Trp (CCA)"
            194..196
            /note="transfer RNA-Trp (CCA) anticodon"
            247..259
            /note="inverted repeat B"
            247..273
            /note="pot.-stem-loop structure; terminator-like structure"
            261..273
            /note="inverted repeat B"
BASE COUNT      80 a 50 c 57 g 94 t
ORIGIN

```

```

Query Match
Best Local Similarity 86.4%; Pred. No. 1.3e+02;
Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGGAATAGGC 22
Db 74 TGTCTACCTACTGGAATATGC 53

RESULT 8
MCTRW2/c
LOCUS      MCTRW2      76 bp      mRNA      BCT      31-MAR-1992
DEFINITION Mycoplasma capricolum transfer RNA-Trp (UCA).
ACCESSION  X16767 X07953
NID        g44265

```



```

repeat_region complement(52991..53050)
/note="MIR repeat: matches 234..175 of consensus"
repeat_region complement(60954..60997)
/note="MER34 repeat: matches 524..480 of consensus"
repeat_region complement(61036..61200)
/note="MER39 repeat: matches 409..251 of consensus"
repeat_region 61430..61537
/note="MIR2 repeat: matches 32..142 of consensus"
repeat_region complement(62029..62337)
/note="AluJo repeat: matches 301..6 of consensus"
repeat_region 62452..62815
/note="MER39 repeat: matches 48..409 of consensus"
repeat_region 62854..62898
/note="MER34 repeat: matches 480..525 of consensus"
repeat_region 63567..65020
/note="SVA repeat: matches 4..1372 of consensus"
misc_feature complement(65545..65881)
/note="match: STS G05436"
repeat_region complement(65917..66232)
/note="MER2 repeat: matches 345..9 of consensus"
repeat_region complement(67491..67789)
/note="AluJb repeat: matches 300..1 of consensus"
repeat_region complement(68357..68545)
/note="L1 repeat: matches 3768..3579 of consensus"
prim_transcript complement(70789..71348)
/feature="dJ501N12.3"
/note="possibly a pseudogene; match: cDNAs Y10202 283943"
/evidence=not_experimental

```

```

Query Match      83.6%; Score 18.4; DB 9; Length 170952;
Best Local Similarity 95.0%; Pred. No. 20;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTCTACCACTGAACCTAG 20
    ||| ||||| ||||| |||||
Db 38328 TGCTCTACCACTGAACCTAG 38309

```

```

RESULT 4
ALCRDNA/c
LOCUS          ALCRDNA      6081 bp      DNA      PLN      31-MAR-1992
DEFINITION     Astasia longa chloroplast ribosomal DNA.
ACCESSION      X14386
NID            g11194
VERSION        X14386.1 GI:111194
KEYWORDS       16S ribosomal RNA; 23S ribosomal RNA; 5S ribosomal RNA; ribosomal
               DNA; ribosomal RNA; transfer RNA; transfer RNA-val.
SOURCE          euglenophyceae alga.
ORGANISM        Chloroplast Astasia longa
REFERENCE       Eukaryota; Euglenozoa; Euglenida; Euglenales; Astasia.
AUTHORS         Siemeister,G.
TITLE           Direct Submission
JOURNAL         Submitted (08-FEB-1989) Siemeister G., Botanisches Institut der
               Universiteit, Kirschallee 1, D 5300 Bonn 1, FRG
REFERENCE       2. (bases 1 to 6081)
AUTHORS         Siemeister,G. and Hachtel,W.
TITLE           Organization and nucleotide sequence of ribosomal RNA genes on a
               circular 73 kbp DNA from the colourless flagellate Astasia longa
JOURNAL         Curr. Genet. 17 (5), 433-438 (1990)
MEDLINE         90291517
FEATURES        Location/Qualifiers
                 source      1..6081
                        /organism="Astasia longa"
                        /chloroplast
                        /strain="CCAP 1204-17a"
                        /db_xref="taxon:3037"
                        34..3137
                        /note="23S ribosomal RNA"
                        3252..3382
                        /note="5S ribosomal RNA"
                        3674..3746
                        /note="trna-val"

```

```

rRNA          3860..5379
/note="16S ribosomal RNA"
rRNA          5535..>6081
/note="23S ribosomal RNA"
BASE COUNT    2040 a 723 c 1301 g 2017 t
ORIGIN

Query Match    79.1%; Score 17.4; DB 7; Length 6081;
Best Local Similarity 94.7%; Pred. No. 81;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGTCTACCACTGAACCTA 19
    ||||| ||||| ||||| |||||
Db 3699 TGTTCTACCACTTGAACCTA 3681

RESULT 5
ECOTRW/c
LOCUS          ECOTRW      76 bp      tRNA      BCT      20-MAY-1994
DEFINITION     E.coli Trp-tRNA; also ts and uga-suppressor mutations.
ACCESSION      K00260
NID            g174464
VERSION        K00260.1 GI:174464
KEYWORDS       transfer RNA; transfer RNA suppressor; transfer RNA-Trp.
SOURCE         Escherichia coli ([1]: wild-type strain CA244, uga-suppressor
               strain caJ64; [2]: wild-type strain LS 340, temperature-sensitive
               strain LS 874) tRNA.
ORGANISM        Escherichia coli
               Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
               Escherichia.
REFERENCE       1. (bases 1 to 76)
AUTHORS         Hirsh,D.
TITLE           Tryptophan transfer RNA as the UGA suppressor
JOURNAL         J. Mol. Biol. 58, 439-458 (1971)
MEDLINE        71234596
REFERENCE       2. (bases 1 to 76)
AUTHORS         Eisenberg,S.P., Soell,L. and Yarus,M.
TITLE           The purification and sequence of a temperature-sensitive tryptophan
               tRNA
JOURNAL         J. Biol. Chem. 254, 5562-5566 (1979)
MEDLINE        79194094
COMMENT        Contributed on tape April 1983 by M.Sprinzi & D.H.Gauss: from their
               entries 1810, 1811, 1814 in Nucleic Acids Res. 11, r1-r54 (1983).
               [1] suggests that 'uga' suppression is carried out by a Trp-tRNA
               whose anticodon-codon recognition is altered by a change outside
               the anticodon. [2] discusses the mutational effects on the in vitro
               and in vivo properties of Trp-tRNA.
FEATURES        Location/Qualifiers
                 source      1..76
                        /organism="Escherichia coli"
                        /db_xref="taxon:562"
                        1..76
                        /note="codon recognized: TGG; Trp-tRNA (NAR: 1810)"
                        /product="trna-Trp"
                 mutation    7
                        /notes="g in wild-type; a in ts mutant ls 874 [2]"
                 modified_base 8
                        /mod_base=s4u
                 modified_base 16
                        /mod_base=d
                 modified_base 17
                        /mod_base=d
                 modified_base 20
                        /mod_base=d
                 modified_base 24
                        /mod_base=d
                 mutation    24
                        /note="a in wt (su-uga); g in suppressor mutant su-uga
                        [1]"
                 modified_base 32
                        /mod_base=cm
                 modified_base 37
                        /mod_base=ms216a
                 modified_base 46

```

only a small overlap as described above. This sequence is the entire insert of clone 501N12. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed in collaboration by the Sanger Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr6/>

501N12 is from the library RPCI3 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

FEATURES

Source

1. .170952
 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="6"
 /clone="501N12"
 /map="p22.1-22.3"
 /clone_lib="RPCI3"
 444. .516

repeat_region
 /note="MER5A repeat: matches 101. .188 of consensus"
 complement(832. .1100)
 repeat_region
 /note="AluJb repeat: matches 280. .1 of consensus"
 complement(1341. .1477)
 repeat_region
 /note="MIR repeat: matches 237. .85 of consensus"
 2942. .3479
 repeat_region
 /note="L1 repeat: matches 1432. .1968 of consensus"
 4030. .4359
 repeat_region
 /note="AluY repeat: matches 1. .296 of consensus"
 complement(4554. .4853)
 repeat_region
 /note="AluJo repeat: matches 301. .1 of consensus"
 6829. .34852
 gene
 /gene="dJ501N12.1"
 CDS
 join(-6829. .6895,15857. .15985,17949. .18080,28912. .29087,32011. .32098,34703. .>34852)
 /gene="dJ501N12.1"
 /note="match: cDNAs Y10519 Y10519 283950 Y10508"
 /codon_start=2
 /evidence="not_experimental"
 /product="dJ501N12.1"
 /protein_id="CAA18156.1"
 /db_xref="PID:el312796"
 /db_xref="PID:g3355565"
 /db_xref="GI:3355565"
 /db_xref="SPTREMBL:O75660"
 /translation="RKYLEQAYRLQOGIVTSTQOMIDRICVKVDHLNLSRNGGDA
 ATQEDLKSERLNRDKNKNTLLPNYHVGGSWAGSLSSPQETLESWAGEVTR
 VDEQKALLESMDAENLCNPMKKAHQRDLTHASTEKISIPRTFVKYNVLLQSG
 IDILNKISVKLTVAFLSDRIVDEILDALSHKHKLADHFSRRGKTLTPQESLEIEL
 AEEKPKVRSIIIVTEELTEIERLEDLTCM"
 prim_transcript
 7271. .7658
 /gene="dJ501N12.1"
 /note="match: EST AA526622"
 repeat_region
 7659. .7776
 /note="AluYb repeat: matches 191. .308 of consensus"
 repeat_region
 8239. .8475
 /note="AluSg repeat: matches 1. .239 of consensus"
 complement(8536. .8952)
 misc_feature
 /gene="dJ501N12.1"
 /note="match: GSS B37778"
 repeat_region
 complement(9800. .10072)
 /note="AluX repeat: matches 273. .1 of consensus"
 repeat_region
 complement(10099. .10222)
 /note="AluSp repeat: matches 303. .179 of consensus"
 repeat_region
 11294. .11503
 /note="MIR repeat: matches 4. .214 of consensus"
 prim_transcript
 complement(12237. .12525)
 /gene="dJ501N12.1"
 /note="match: EST AA639318"

repeat_region
 complement(12599. .12717)
 /note="MIR2 repeat: matches 146. .28 of consensus"
 15448. .15519
 repeat_region
 /note="MIR2 repeat: matches 74. .146 of consensus"
 complement(15866. .16333)
 misc_feature
 /gene="dJ501N12.1"
 /note="match: GSS AQ016761"
 complement(16512. .16606)
 repeat_region
 /note="MER21B repeat: matches 790. .703 of consensus"
 complement(16545. .16637)
 repeat_region
 /note="MER39 repeat: matches 676. .582 of consensus"
 complement(16603. .17427)
 repeat_region
 /note="MER21B repeat: matches 792. .4 of consensus"
 17480. .17647
 repeat_region
 /note="AluSp repeat: matches 136. .303 of consensus"
 complement(19382. .19433)
 repeat_region
 /note="L1HS repeat: matches 578. .526 of consensus"
 complement(19656. .19960)
 repeat_region
 /note="AluSx repeat: matches 300. .1 of consensus"
 22609. .22648
 repeat_region
 /note="20 copies 2 mer ac 100% conserved"
 complement(22914. .23103)
 repeat_region
 /note="MER3 repeat: matches 209. .1 of consensus"
 24161. .24346
 repeat_region
 /note="L1 repeat: matches 4042. .4231 of consensus"
 24420. .24689
 repeat_region
 /note="AluJo repeat: matches 1. .302 of consensus"
 24771. .25083
 repeat_region
 /note="AluSg repeat: matches 2. .303 of consensus"
 26805. .26904
 repeat_region
 /note="MIR repeat: matches 35. .134 of consensus"
 complement(27877. .28054)
 repeat_region
 /note="MIR repeat: matches 260. .77 of consensus"
 28097. .28238
 repeat_region
 /note="MIR repeat: matches 71. .201 of consensus"
 complement(28312. .28688)
 misc_feature
 /gene="dJ501N12.1"
 /note="match: GSS AQ020591"
 complement(30856. .30913)
 repeat_region
 /note="MER5A repeat: matches 111. .57 of consensus"
 complement(33402. .33698)
 repeat_region
 /note="AluX repeat: matches 301. .1 of consensus"
 complement(38477. .38699)
 repeat_region
 /note="MIR repeat: matches 253. .2 of consensus"
 complement(40249. .40548)
 repeat_region
 /note="AluSg repeat: matches 299. .1 of consensus"
 complement(41636. .41692)
 repeat_region
 /note="MIR2 repeat: matches 146. .87 of consensus"
 42990. .43016
 repeat_region
 /note="MIR2 repeat: matches 98. .124 of consensus"
 43567. .43860
 repeat_region
 /note="AluSc repeat: matches 1. .298 of consensus"
 44628. .44723
 repeat_region
 /note="MIR2 repeat: matches 30. .134 of consensus"
 complement(45009. .45166)
 repeat_region
 /note="MIR repeat: matches 180. .21 of consensus"
 complement(47974. .48264)
 repeat_region
 /note="AluY repeat: matches 301. .2 of consensus"
 complement(48410. .49073)
 mRNA
 /gene="dJ501N12.2"
 /note="match: cDNA Y10305"
 /pseudo
 /evidence="not_experimental"
 complement(48410. .49073)
 gene
 /gene="dJ501N12.2"
 49166. .50348
 repeat_region
 /note="MER42c repeat: matches 348. .1536 of consensus"
 50415. .50724
 repeat_region
 /note="AluJo repeat: matches 2. .302 of consensus"
 complement(50998. .51085)
 repeat_region
 /note="MIR2 repeat: matches 145. .58 of consensus"
 complement(52655. .52797)
 repeat_region
 /note="L1ME2 repeat: matches 699. .554 of consensus"

```

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGAAGTGGC 22
|||||
Db 1 TGTCTACCACTGAAGTGGC 22

RESULT 2
HSII1AG 11970 bp DNA PRI 24-APR-1993
DEFINITION Human gene for interleukin 1 alpha (IL-1 alpha).
ACCESSION X03833
NID 933785
VERSION X03833.1 GI:33785
KEYWORDS Alu repetitive sequence; interleukin 1 alpha; inverted repeat;
repetitive sequence.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Furutani,Y., Notake,M., Fukui,T., Ohue,M., Nomura,H., Yamada,M. and
Nakamura,S.
TITLE Complete nucleotide sequence of the gene for human interleukin 1
alpha
JOURNAL Nucleic Acids Res. 14 (8), 3167-3179 (1986)
MEDLINE 86205226
REMARK Erratum:[published erratum appears in Nucleic Acids Res 1986 Jun
25:14(12):5124]]
COMMENT Data kindly reviewed (10-NOV-1986) by Y. Furutani.
FEATURES
source
location/Qualifiers
1..11970
/organism="Homo sapiens"
/db_xref="taxon:9606"
254..393
/note="Alu repetitive sequence"
misc_feature
1375..1382
/note="direct repeat 1"
repeat_region
1375..1390
/note="pot. transcription activator (seq. homolog. to
adenovirus 2 major late promoter transcription factor
(MLF) binding site)"
repeat_unit
1375..1382
/note="inverted repeat A"
repeat_unit
1383..1390
/note="inverted repeat A'"
repeat_region
1383..1390
/note="direct repeat 1"
TATA_signal
1407..1413
prim_transcript
1438..11643
exon
1438..1488
/number=1
mRNA
join(1438..1488,2153..2207,3166..3214,4103..4325,
6262..6432,7815..7939,10290..11643)
intron
1489..2152
/number=1
exon
2153..2207
/number=2
CDS
join(2151..2207,3166..3214,4103..4325,6262..6432,
7815..7939,10290..10490)
/product="IL-1-alpha"
/protein_id="CAA27448.1"
/db_xref="PID:933786"
/db_xref="GI:33786"
/db_xref="SWISS-PROT:P01583"
/translation="MAKVPDMFEDLNKNCYSENEEDSSIDHLSLNOKSFYHVSXGPLH
EGCQDSVLSISFSETSKLTFKESMVVATNGVKLRRLSLSOSITDDLEAIAN
DSEELIKPRSPFSFLSNVKNFRIKIYEFILNDALNQSIIRANDQYLYLAALHNL
DEAVKFDGAKYSSKDDAKIIVILRSKITQIVTAQDEDPVLKEMPEIKPTITGSE
TNLFFWETHGKTNFTSVAHPNLFIAIKQDIYVWGLAGGPPSITDFQILENQA"
2208..3165
/number=2
intron

```

```

exon
3166..3214
/number=3
intron
3215..4102
/number=3
exon
4103..4325
/number=4
intron
4326..6261
/number=4
misc_feature
4893..5174
/note="Alu repetitive sequence"
exon
6262..6432
/number=5
intron
6433..7814
/number=5
misc_feature
7695..7744
/note="poly [dA-dC] tract"
exon
7815..7939
/number=6
intron
7940..10289
/number=6
misc_feature
8466..8483
/note="poly[dA-dC] tract"
repeat_region
8912..9137
/note="5 x 46 bp repeat"
misc_feature
9770..9806
/note="poly [dT-dG] tract"
exon
10290..11643
/number=7
polyA_site
11643
/note="polyA site"
misc_feature
11863..11970
/note="Alu repetitive sequence"
BASE COUNT 3708 a 2489 c 2226 g 3547 t
ORIGIN

```

```

Query Match 100.0%; Score 22; DB 10; Length 11970;
Best Local Similarity 100.0%; Pred. No. 0.42;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGTCTACCACTGAAGTGGC 22
|||||
Db 471 TGTCTACCACTGAAGTGGC 492

```

```

RESULT 3
LOCUS HS501N12/c DNA PRI 24-JUL-1998
DEFINITION HS501N12 170952 bp Homo sapiens DNA sequence from clone 501N12 on chromosome
6p22.1-22.3. Contains a gene almost identical to four genes of
unknown function, a pseudogene, three (pseudo?) genes similar to
genes of unknown function, an unknown gene similar to a rat EST, a
PX19 LIKE pseudogene and another unknown gene. Contains ESTs, STSS
and GSSs, complete sequence.
ACCESSION AL022170
NID 93281976
VERSION AL022170.1 GI:3281976
KEYWORDS HTG; PX19.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS 1 (bases 1 to 170952)
TITLE Direct Submission
JOURNAL Submitted (06-JUL-1998) E-mail enquires: humquery@sanger.ac.uk
COMMENT
Clone requests: clonerequest@sanger.ac.uk
On Jul 1, 1998 this sequence version replaced gi:2980811.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with

```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.
OM nucleic - nucleic search, using sw model
Run on: September 18, 1999, 15:49:15 ; Search time 436.05 Seconds
(without alignments)
160.456 Million cell updates/sec

Title: US-09-037-472-7
Perfect score: 22
Sequence: 1 TGTCTACCACTGAAGTAGGC 22

Scoring table: IDENTITY_NUC
Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*
1: gb_bal.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sts.*
15: gb_sy.*
16: gb_un.*
17: gb_v1.*
18: em_fun.*
19: em_htg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_om.*
24: em_or.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_ro.*
30: em_sts.*
31: em_sy.*
32: em_un.*
33: em_v1.*
34: gb_hgt1.*
35: gb_hgt2.*
36: gb_in1.*
37: gb_in2.*
38: em_bal.*
39: em_ba2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	22	100.0	22	5	173225 Sequence 1

2	100.0	11970	10	HSIL1AG
3	83.6	170952	9	HS501N12
4	79.1	6081	7	ALCRDNA
5	78.2	76	1	ECOTRW
6	78.2	76	1	ECOTRWSUP
7	78.2	281	1	MCTRNW12
8	78.2	76	1	MCTRNW2
9	78.2	994	1	MYCTGW
10	78.2	300	1	MYCTRGCT
11	78.2	39897	34	LMFL1156
12	78.2	43511	34	LMFL1684
13	76.4	135551	1	D90906
14	76.4	1895	1	SPMULTG
15	76.4	119704	7	OSCHLPLXX
16	76.4	8829	11	U73166
17	76.4	110000	34	AC005079_0
18	76.4	58780	35	AC005085
19	76.4	85096	35	AC007323
20	74.5	13555	2	AE000670
21	74.5	150613	7	AB001684
22	74.5	135599	7	CPU30821
23	74.5	121524	8	AF041468
24	74.5	191028	8	PPU38804
25	74.5	55328	8	PMU02970
26	74.5	201986	12	AC006289
27	74.5	201964	12	MMHCC29N7
28	74.5	252522	34	HSJ874H6
29	74.5	138894	34	HSJ878113
30	74.5	165321	35	AC006549
31	73.6	36368	1	SC9B5
32	73.6	15580	2	AE000776
33	73.6	11824	2	AE001155
34	73.6	1161	3	AF028000
35	73.6	2682	3	GOTMTGRG
36	73.6	28199	4	AF013613
37	73.6	55892	4	AF013614
38	73.6	16783	4	AF106038
39	73.6	82697	7	ATT9A21
40	73.6	3967	7	BOSRK29G
41	73.6	3057	8	AF038122
42	73.6	107603	9	HS934G17
43	73.6	103574	10	HSAC002115
44	73.6	4211	10	HSDRES9
45	73.6	48588	35	AC007581

ALIGNMENTS

RESULT 1	173225	Sequence 1 from patent US 5686246.	22 bp	DNA	PAT	23-DEC-1997
LOCUS	I73225					
DEFINITION						
ACCESSION	I73225					
NID	G3009364					
VERSION	I73225.1	GI:3009364				
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 22)					
AUTHORS	Korman,K.S. and Duff,G.W.					
TITLE	Detecting genetic predisposition to periodontal disease					
JOURNAL	Patent: US 5686246-A 11-NOV-1997;					
FEATURES	Location/Qualifiers					
Source	1..22					
BASE COUNT	5 a	/organism="unknown"	7 c	4 g	6 t	
ORIGIN						

Query Match 100.0%; Score 22; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. NO. 0.69;

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 15:49:01 ; Search time 436.05 Seconds
(without alignments)
123.989 Million cell updates/sec

Title: US-09-037-472-6

Perfect score:

Sequence: 1 TCCTGGTCTGCAGGTAA 17

Scoring table: IDENTITY NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl:★

```

1: gb_bal.*
2: gb_ba3.*
3: gb_Om.*
4: gb_Ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pll.*
8: gb_pl2.*
9: gb_pri.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sts.*
15: gb_sy.*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
No.						
1	17	100.0	17	5	I85579	Sequence 2

C	2	17	100.0	12565	10	HS1IL1RCA
C	3	16	94.1	186510	9	HS451LB15
C	4	16	94.1	113328	34	HSJ219P5
C	5	15.4	90.6	371170	4	FR258N02
C	6	15.4	90.6	121780	9	HS774I24
C	7	15.4	90.6	132953	11	AC003015
C	8	15.4	90.6	119483	11	AC005588
C	9	15.4	90.6	33414	11	HS065590
C	10	15.4	90.6	159137	34	CNS000007
C	11	15.4	90.6	181832	34	HSJ1069C8
C	12	15.4	90.6	91047	35	AC007491
C	13	15	88.2	13143	2	U67563
C	14	15	88.2	21935	4	AB016081
C	15	15	88.2	122325	9	HS1044017
C	16	15	88.2	130926	9	HS453A3
C	17	15	88.2	115743	9	HS488H17
C	18	15	88.2	128169	9	HS849L7
C	19	15	88.2	116210	10	HS4066C18
C	20	15	88.2	101046	10	HS936P19
C	21	15	88.2	106000	10	HSU40455
C	22	15	88.2	188205	11	AC003986
C	23	15	88.2	152798	11	AC004057
C	24	15	88.2	207957	11	AC004470
C	25	15	88.2	128559	11	AC004577
C	26	15	88.2	164511	11	AC004894
C	27	15	88.2	196361	11	AC005386
C	28	15	88.2	185116	11	AC005823
C	29	15	88.2	177707	11	AC006365
C	30	15	88.2	5809	12	D89657S4
C	31	15	88.2	4400	12	MM067925
C	32	15	88.2	412	14	G34073
C	33	15	88.2	186682	34	AC005050
C	34	15	88.2	81897	34	AC005916
C	35	15	88.2	203407	34	AC006174
C	36	15	88.2	186820	34	AC006400
C	37	15	88.2	211727	34	CNS00000C
C	38	15	88.2	116835	34	HS676313
C	39	15	88.2	64225	34	HSV237C10.4
C	40	15	88.2	110000	34	HSY313F4.0
C	41	15	88.2	110000	34	HSY738F9_0
C	42	15	88.2	211522	35	AC005883
C	43	15	88.2	150342	35	AC005973
C	44	15	88.2	190653	35	AC006994
C	45	15	88.2	176258	42	AC007102

ALIGNMENTS

RESULT	1
185579	
LOCUS	I85579
DEFINITION	Sequence 2 from patent US 5698399.
ACCESSION	I85579
NID	G3205297
VERSION	I85579.1 GI:3205297
KEYWORDS	-
SOURCE	Unknown.
ORGANISM	Unknown.
REFERENCE	Unclassified.
AUTHORS	1 (bases 1 to 17)
TITLE	Duff, G.W., Russell, G. and Eastell,
JOURNAL	Detecting genetic predisposition for
FEATURES	Patent: US 5698399-A 16-DEC-1997
source	Location/Qualifiers 1..17 /organism="unknown"
BASE COUNT	3 a 4 c 5 g
ORIGIN	5

Query Match	100.0%	Score 17;	DB 5;	Length 17;
Best Local Similarity	100.0%	pred. No. 7.9;		


```
repeat_region 19891..20513
/Note="L1 repeat: matches 5390. .4746 of consensus"
25324..28347
/Gen="dJ451B15.2"
CDS join(<25324..25815,27169..28347)
/Gen="dJ451B15.2"
/codon_start=1
/product="DNA binding protein"
/protein_id="CAB10847.1"
/db_xref="PID:e1237440"
/db_xref="PID:g2791274"
/db_xref="GI:2791274"
/db_xref="SPTREMBL:O43733"
/translation="DEKORFSYRSGYDLESDGDPEDDNEDEDDSQAESVLSAT
PSYATSPQLPSRSLQDPVSTDEDRITDFSGVHTDPMVLPRALITRMVLSTAQ
SDYKNTLSPKQRAARDNDITPSVDTSRPCHQMSVDYPESEETLRSSMAGKAV
AITQSPSVRLPPAAAEHSPOTAAAGMPSVASPHDPQKQOJTLQPTPGLPSPHTHL
FSLPLHSQOQSRLPYNMVPGGHHVVPAGLITFTFPLQAGPQLIIPAVSVHRTL
GTHRTVTESGTTNPAGVAELSSVPCIPIGQIRVPGQLNSTPGQLSPLSMETH
NIVGLANTMAQVHPHPPGALNAGVLQVLTANPSSQSPAPQAHIPGLIINIALPTL
IPSVQAVDAQAGAPASQSKACETQPKTQSVASANQSVRTSPQGLPTVQRENK
KVLNPPAPAGHARDLGLSKMDTEKAASAHVKKPELTSIQOGPASTSQPLLKAHSE
VFTKPSGOQTLSPRQVPRPTALPRQPTVHFSDVSDDDDEDLVIAT"
31502..31558
/Note="3 copies of 19 mer 84 % conserved"
31543..31572
/Note="15 copies of 2 mer 87 % conserved"
31821..31852
/Note="16 copies of 2 mer 84 % conserved"
31854..31885
/Note="16 copies of 2 mer 84 % conserved"
31887..31918
/Note="16 copies of 2 mer 84 % conserved"
31920..31999
/Note="40 copies of 2 mer 83 % conserved"
32774..33070
/Note="AlusX repeat: matches 302. .3 of consensus"
33746..34022
/Note="AluY repeat: matches 22. .297 of consensus;
incomplete repeat"
35106..35237
/Note="AluJ repeat: matches 133. .1 of consensus;
incomplete repeat"
38680..38706
/Note="9 copies of 3 mer 93 % conserved"
39199..39348
/Note="MIR repeat: matches 96. .255 of consensus"
40120..40417
/Note="AlusQ repeat: matches 300. .1 of consensus"
40566..40763
/Note="L1 repeat: matches 5296. .5092 of consensus"
40764..41213
/Note="L1PA15 repeat: matches 45. .904 of consensus"
41242..41471
/Note="MIR repeat: matches 10. .244 of consensus"
42195..42316
/Note="AluJ repeat: matches 3. .124 of consensus;
incomplete repeat"
42335..42802
/Note="L1MB3 repeat: matches 85. .565 of consensus"
43956..44097
/Note="MIR repeat: matches 243. .87 of consensus"
44966..45265
/Note="AluY repeat: matches 2. .301 of consensus"
45286..45459
/Note="AluJ repeat: matches 290. .138 of consensus;
incomplete repeat"
47861..47952
/Note="23 copies of 4 mer 88 % conserved"
49463..49981
/Note="L1MB7 repeat: matches 30. .547 of consensus"
49993..50122
/Note="L1MB8 repeat: matches 457. .586 of consensus"
```

```
repeat_region 50176..50432
/Note="L1MB8 repeat: matches 589. .866 of consensus"
51118..51414
/Note="MSTB repeat: matches 1. .289 of consensus"
51478..51771
/Note="AlusX repeat: matches 301. .1 of consensus"
51772..51840
/Note="MSTC repeat: matches 335. .403 of consensus"
53562..53808
/Note="L1MB3 repeat: matches 296. .551 of consensus"
55194..55497
/Note="AluJ repeat: matches 1. .301 of consensus"
55790..55887
/Note="MIR repeat: matches 157. .58 of consensus"
56188..56323
/Note="L1MA44 repeat: matches 1043. .901 of consensus"
56324..56703
/Note="MSTA repeat: matches 1. .426 of consensus"
56845..56952
/Note="MIR repeat: matches 66. .186 of consensus"
56953..57025
/Note="AlusX/g repeat: matches 186. .121 of consensus;
incomplete repeat"
56953..57003
/Note="AluSp/q repeat: matches 51. .1 of consensus;
incomplete repeat"
57113..57518
/Note="L1MD1 repeat: matches 537. .119 of consensus"
57891..58154
/Note="MER33 repeat: matches 26. .303 of consensus"
58246..58269
/Note="12 copies of 2 mer 96 % conserved"
58405..58707
/Note="AlusX repeat: matches 1. .302 of consensus"
59100..59223
/Note="MIR2 repeat: matches 21. .145 of consensus"
59379..59542
/Note="MIR repeat: matches 237. .58 of consensus"
60764..60867
/Note="MIR2 repeat: matches 144. .41 of consensus"
61162..61456
/Note="AlusX repeat: matches 1. .300 of consensus"
62408..62451
/Note="MIR2 repeat: matches 56. .15 of consensus"
62689..62801
/Note="MLT2FB repeat: matches 146. .1 of consensus"
63167..63214
/Note="MSTA repeat: matches 376. .423 of consensus"
63247..63536
/Note="AlusQ repeat: matches 301. .10 of consensus"
65081..65380
/Note="AluJb repeat: matches 1. .301 of consensus"
66041..66324
/Note="AluJb repeat: matches 301. .1 of consensus"
67681..67979
/Note="AlusQ repeat: matches 1. .300 of consensus"
68349..68635
/Note="MIR2 repeat: matches 146. .47 of consensus"
69076..69370
/Note="AluY repeat: matches 1. .295 of consensus"
69400..69589
/Note="MIR repeat: matches 1. .213 of consensus"
```

Query Match 94.1%; Score 16; DB 9; Length 186510;
Best Local Similarity 100.0%; Pred.No.31;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCTGGTCTGCAGGTAA 17

Db 166415 CCTGGTCTGCAGGTAA 166400

RESULT 4


```
exon      complement(23376. .>23450)
          /gene="IGF2"
          /number=1
mRNA      complement(join(<25186. .25345,25675. .25808,25886. .25981,
26359. .26415,26499. .26568,26668. .26803,26914. .27059,
27421. .27471,27552. .27619,27753. .27841,27943. .28117,
28323. .28526,28846. .>28929))
          /gene="TH"
          /number=5
gene      complement(25186. .28929)
          /gene="TH"
exon      complement(<25186. .25345)
          /gene="TH"
          /number=13
CDS       complement(join(25186. .25345,25675. .25808,25886. .25981,
26359. .26415,26499. .26568,26668. .26803,26914. .27059,
27421. .27471,27552. .27619,27753. .27841,27943. .28117,
28323. .28526,28846. .28929))
          /gene="TH"
          /codon_start=1
          /product="tyrosine 3-monooxygenase"
          /protein_id="CAAL17124.1"
          /db_xref="PID:e1251684"
          /db_xref="PID:q2894578"
          /db_xref="GI:2894578"
          /translation="MPQSGSTSTKSTRRAPSELERSDSVTSQFLGRROSLLIEDAR
KERAAAAAEEAEEAEEQIVFEEDDGRALLNFFLRNTKTPALSRILKRVFETFEAKI
HHELRPCPKLKSLEGLYFVRCLEHLSDVSLIGSLKRNAEDVKTKEVFKHFWPK
KIADLKCHLVTKFDPLDQDPGYTDAAYRQRKMIGDIAFRVNGEPIPRVEYTE
EEIGTWREYVLTLDLYLTHACSEHLEAFRLLEKHGYSPONIPOLEDVSYFLKEHTG
FILRPVAGILLNARDLASLAFRFVQCTQYIRHASSPMHSPEDCVHELLGHVPLMADS
TFAQSQSLGASGASDEDEIKLSTLYFTVEYGLCKONGEVKAYGAGLLSSYGELY
HSLSDPEVREDFDPAAYQVQDTYQPVYFISESFADAKEKFRYYVAGIKRPFPSVR
FDPYTTSIQVLNPLKIQGLECVKDELKMLADALSVLs"
          /complement(25344. .25676)
          /gene="TH"
          /number=12
exon      complement(25675. .25808)
          /gene="TH"
          /number=12
intron    complement(25807. .25887)
          /gene="TH"
          /number=11
exon      complement(25886. .25981)
          /gene="TH"
          /number=11
intron    complement(25980. .26360)
          /gene="TH"
          /number=10
exon      complement(26359. .26415)
          /gene="TH"
          /number=10
intron    complement(26414. .26500)
          /gene="TH"
          /number=9
exon      complement(26499. .26568)
          /gene="TH"
          /number=9
intron    complement(26567. .26669)
          /gene="TH"
          /number=8
exon      complement(26668. .26803)
          /gene="TH"
          /number=8
intron    complement(26802. .26915)
          /gene="TH"
          /number=7
exon      complement(26914. .27059)
          /gene="TH"
          /number=7
intron    complement(27058. .27422)
          /gene="TH"
          /number=6
exon      complement(27421. .27471)
```

```
          /gene="TH"
          /number=6
          complement(27470. .27553)
          /gene="TH"
          /number=5
          complement(27552. .27619)
          /gene="TH"
          /number=5
          complement(27618. .27754)
          /gene="TH"
          /number=4
          complement(27753. .27841)
          /gene="TH"
          /number=4
          complement(27840. .27944)
          /gene="TH"
          /number=3
          complement(27943. .28117)
          /gene="TH"
          /number=3
          complement(28116. .28324)
          /gene="TH"
          /number=2
          complement(28323. .28526)
          /gene="TH"
          /number=2
          complement(28525. .28847)
          /gene="TH"
          /number=1
          complement(28846. .>28929)
          /gene="TH"
          /number=1
          repeat_region
          30492..30543
          /note="CA"
          /rpt_type=TANDEM
          join(<36343. .36352,36516. .36588,36683. .>36831)
          /gene="NAP2"
          36343..36831
          /gene="NAP2"
```

Query Match 90.6%; Score 15.4; DB 4; Length 37170;
Best Local Similarity 94.1%; Pred. No. 70;

Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTGGTCGCAGGTAA 17
||||| |||||
Db 23205 TCCTGGTTGCAGGTAA 23189

RESULT 6

HS774124

LOCUS

DEFINITION

HS774124 121780 bp DNA PRI 24-NOV-1998
Human DNA sequence from clone 774124 on chromosome 1q24.1-24.3
Contains protein similar to pregnancy-associated plasma protein A
precursor neuronal migration protein astrotactin, ESTs, STS and
GSS, complete sequence.

ACCESSION

AL031290

NID

g3550114

VERSION

AL031290.1 GI:3550114

KEYWORDS

HTG; neuronal migration protein astrotactin; pregnancy-associated

plasma protein A precursor.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 121780)

Grafham,D.

Direct Submission

Submitted (06-OCT-1998) Sanger Centre, Hinxton, Cambridgeshire,

CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone

requests: clonerequest@sanger.ac.uk

On Sep 8, 1998 this sequence version replaced gi:3421066.

During sequence assembly data is compared from overlapping clones.


```

repeat_region /note="MER5A repeat: matches 14. .167 of consensus"
complement(33181. .34085)
repeat_region /note="LIPAI2 repeat: matches 911. .1 of consensus"
complement(33940. .35470)
repeat_region /note="L1 repeat: matches 5390. .3854 of consensus"
complement(35944. .36375)
repeat_region /note="MER4A repeat: matches 661. .229 of consensus"
complement(36372. .36548)
repeat_region /note="MER4A repeat: matches 173. .1 of consensus"
36549. .36620
repeat_region /note="3 copies 24 mer 88% conserved"
36552. .36619
repeat_region /note="34 copies 2 mer ta 90% conserved"
36773. .36810
repeat_region /note="19 copies 2 mer ca 92% conserved"
38127. .38156
repeat_region /note="15 copies 2 mer aa 87% conserved"
39389. .39534
repeat_region /note="MIR2 repeat: matches 2. .144 of consensus"
40530. .40559
repeat_region /note="6 copies 5 mer aaac 93% conserved"
complement(40965. .41320)
misc_feature /gene="d3774124.1"
complement(41199. .41238)
repeat_region /note="match: AL009861 Chromosome 1 HindIII fragment"
41199. .41238
repeat_region /note="20 copies 2 mer aa 80% conserved"
45085. .45199
repeat_region /note="MIR repeat: matches 65. .187 of consensus"
complement(46396. .46552)
repeat_region /note="MIR repeat: matches 172. .19 of consensus"
48619. .48678
repeat_region /note="MIR repeat: matches 124. .135 of consensus"

```

Query Match 90.6%; Score 15.4; DB 9; Length 121780;
 Best Local Similarity 94.1%; Pred No. 70;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCGTGGCTGCAGGTAA 17
 Db 111475 TCGTGGCTGCAGGTAA 111491
 ||||| |||||

```

RESULT 7
AC003015/c AC003015 132953 bp DNA PRI 20-OCT-1997
DEFINITION Human BAC clone GS113H23 from 5p15.2, complete sequence.
ACCESSION AC003015
NID 92547255
VERSION AC003015.1 GI:2547255
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 132953)
AUTHORS Wamsley,P, Kramer,J, Elliott,G and O'Brien,D.
TITLE The sequence of H. sapiens BAC clone GS113H23
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 132953)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (20-OCT-1997) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT SUBMITTED BY:
Genome Sequencing Center
Department of Genetics
Washington University
St. Louis MO 63108, USA
http://genome.wustl.edu/gsc
mailto:sapiens@watson.wustl.edu

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University School of Medicine, St. Louis MO.

Mapping information for this clone was also provided by Dr. Michael Lovett, Departments of Otorhinolaryngology, Molecular Biology and Oncology, University of Texas Southwestern Medical Center, Dallas TX.

SOURCE INFORMATION:

This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).

Cell line: lymphoblastoid

Haplotypes: two

VECTOR: pBelOBAC

Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is GS330J10, 200 bp overlap. Actual start of this clone is at base position 1 of GS113H23; actual end is at 132953 of GS113H23.

This clone contains STS HSC022YA5 (NID:g1235481) and HS2682D9 (NID:g454585).

This clone contains polymorphisms with GS330J10.

FEATURES

Source	Location/Qualifiers
	1. .132953
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="5"
	/clone="GS113H23"
	/clone_lib="GSBAC1"
	/map="5p15.2"
repeat_region	complement(20. .194)
repeat_region	/rpt_family="L1"
	195. .358
repeat_region	/rpt_family="ALU"
	complement(359. .853)
repeat_region	/rpt_family="L1"
	2903. .2933
repeat_region	/rpt_family="L1"
	complement(3640. .3667)
repeat_region	/rpt_family="L1"
	complement(3764. .3788)
repeat_region	/rpt_family="L1"
	complement(3820. .4112)
repeat_region	/rpt_family="ALU"
	complement(4322. .4351)
repeat_region	/rpt_family="L1"
	complement(4444. .4464)
repeat_region	/rpt_family="L1"
	complement(5046. .5122)
repeat_region	/rpt_family="L1"
	complement(5527. .6871)
repeat_region	/rpt_family="L1"
	complement(6872. .6983)
repeat_region	/rpt_family="ALU"
	complement(6985. .7389)
repeat_region	/rpt_family="L1"
	complement(7459. .8612)
repeat_region	/rpt_family="L1"

```
repeat_region 8604. 9143 /rpt_family="L1"
repeat_region 9170. 9458 /rpt_family="ALU"
repeat_region 9481. 9546 /rpt_family="L1"
repeat_region /rpt_family="L1"
repeat_region complement(9591. 9823)
repeat_region /rpt_family="L1"
repeat_region complement(9854. 9961)
repeat_region /rpt_family="L1"
repeat_region 11262. 11306 /rpt_family="L1"
repeat_region complement(11563. 14264)
repeat_region /rpt_family="L1"
repeat_region complement(14607. 17154)
repeat_region /rpt_family="L1"
repeat_region complement(17883. 17902)
repeat_region /rpt_family="L1"
repeat_region complement(21565. 21588)
repeat_region /rpt_family="L1"
repeat_region 24509. 24814 /rpt_family="ALU"
repeat_region complement(26967. 27276)
repeat_region /rpt_family="L1"
repeat_region 29009. 29051 /rpt_family="L1"
repeat_region complement(29140. 29165)
repeat_region /rpt_family="L1"
repeat_region complement(29892. 30124)
repeat_region /rpt_family="L1"
repeat_region complement(30225. 30258)
repeat_region /rpt_family="L1"
repeat_region 31191. 31565 /rpt_family="L1"
repeat_region 31849. 32016 /rpt_family="L1"
repeat_region 33991. 34027 /rpt_family="L1"
repeat_region 35156. 35179 /rpt_family="L1"
repeat_region /rpt_family="L1"
repeat_region complement(36832. 36880)
repeat_region /rpt_family="L1"
repeat_region 39878. 39924 /rpt_family="L1"
repeat_region complement(41297. 41751)
repeat_region /rpt_family="MER"
repeat_region complement(4442. 44621)
repeat_region /rpt_family="MER"
repeat_region complement(44871. 44963)
repeat_region /rpt_family="MER"
repeat_region complement(45153. 45205)
repeat_region /rpt_family="MER"
repeat_region complement(45255. 45761)
repeat_region /rpt_family="MER"
repeat_region complement(45830. 45882)
repeat_region /rpt_family="MER"
repeat_region complement(45907. 46229)
repeat_region /rpt_family="MER"
repeat_region 46846. 51590 /rpt_family="L1"
repeat_region 51617. 51660 /rpt_family="L1"
repeat_region 52068. 52252 /rpt_family="MER"
repeat_region 52278. 52459 /rpt_family="MER"
repeat_region /rpt_family="MER"
repeat_region complement(52291. 52521)
repeat_region /rpt_family="MER"
repeat_region 53342. 53635 /rpt_family="ALU"
repeat_region 54273. 54297 /rpt_family="L1"
repeat_region 58186. 58479
```

```
repeat_region /rpt_family="ALU"
complement(58482. 58529)
repeat_region /rpt_family="L1"
58482. 58529
repeat_region /rpt_family="L1"
complement(59606. 59897)
repeat_region /rpt_family="ALU"
61733. 62140
repeat_region /rpt_family="L1"
62686. 62745
repeat_region /rpt_family="L1"
complement(63757. 63785)
repeat_region /rpt_family="L1"
complement(63786. 64077)
repeat_region /rpt_family="ALU"
64353. 64376
repeat_region /rpt_family="L1"
67996. 68283
repeat_region /rpt_family="ALU"
complement(68925. 68998)
repeat_region /rpt_family="L1"
68945. 68993
repeat_region /rpt_family="L1"
70132. 70305
repeat_region /rpt_family="ALU"
```

Query Match 90.68; Score 15.4; DB 11; Length 132953;
Best Local Similarity 94.18; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTGCTCTGCAGGTAA 17

Db 34935 TACTGCTCTGCAGGTAA 34919

RESULT 8

AC005588/c AC005588 119483 bp DNA PRI 13-JAN-1999
LOCUS HOMO sapiens PAC clone DJ1161G23 from 7q36, complete sequence.
DEFINITION AC005588
ACCESSION HOMO sapiens PAC clone DJ1161G23 from 7q36, complete sequence.
NID 94153863

VERSION AC005588.1 GI:4153863

KEYWORDS HTG.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 119483)

Sun, H., Bauer, C., Wohldman, P. and Heidbrink, C.

The sequence of Homo sapiens PAC clone DJ1161G23

Unpublished (1999)

REFERENCE 2 (bases 1 to 119483)

Waterston, R.H.

Direct Submission

Submitted (01-SEP-1998) Genome Sequencing Center, Washington

University School of Medicine, 4444 Forest Park Parkway, St. Louis,

MO 63108, USA

3 (bases 1 to 119483)

Waterston, R.

Direct Submission

Submitted (13-JAN-1999) Department of Genetics, Washington

University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

On Jan 13, 1999 this sequence version replaced gi:3907486.

SUBMITTED BY: WUGSC

Genome Sequencing Center

Department of Genetics

Washington University

St. Louis MO 63108, USA

http://genome.wustl.edu/gsc

mailto:sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping

clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.

The clone may be obtained either from Genome Systems, Inc. (<http://www.genomesystems.com>) or Research Genetics, Inc. (<http://www.resgen.com>); or from Pieter de Jong.

VECTOR: pCYPAC2

NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is DJ1048B16. Actual start of this clone is at base position 1 of DJ1161G23; actual end is at 119483 of DJ1161G23.

FEATURES

Source	Location/Qualifiers
	1. .119483
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="7"
	/map="7q36"
	/clone_lib="RPCI-5"
repeat_region	350. .388
	/clone="DJ1161G23"
	/rpt_family="L1"
repeat_region	350. .388
	/rpt_family="5S"
repeat_region	2326. .2375
	/rpt_family="L1"
repeat_region	2415. .7893
	/rpt_family="L1"
repeat_region	7894. .7917
	/rpt_family="L1"
repeat_region	7918. .8860
	/rpt_family="L1"
repeat_region	9207. .9230
	/rpt_family="L1"
repeat_region	10405. .10659
	/rpt_family="L1"
repeat_region	11170. .11190
	/rpt_family="L1"
repeat_region	11324. .11381
	/rpt_family="L1"
repeat_region	12753. .12796
	/rpt_family="L1"
repeat_region	14042. .14141
	/rpt_family="L1"
repeat_region	14312. .14721
	/rpt_family="L1"
repeat_region	14852. .16170
	/rpt_family="L1"
repeat_region	16176. .16746
	/rpt_family="L1"
repeat_region	17029. .17180
	/rpt_family="L1"
repeat_region	17261. .17598
	/rpt_family="L1"
	/rpt_family="L1"
	/note="probable NADH dehydrogenase subunit 6 pseudogene: similar to I80240 (PIB:g2117562)"
misc_feature	17603. .17767
repeat_region	17766. .17963
	/rpt_family="L1"
repeat_region	17975. .18555
	/rpt_family="L1"
repeat_region	19326. .19647
	/rpt_family="L1"
repeat_region	19745. .19813
	/rpt_family="L1"
repeat_region	19842. .19874
	/rpt_family="POLY_A"
repeat_region	20093. .20161
	/rpt_family="L1"
repeat_region	20643. .20714
	/rpt_family="L1"
repeat_region	20716. .20925
	/rpt_family="L1"
repeat_region	21389. .21514
	/rpt_family="L1"
repeat_region	21511. .21636
	/rpt_family="L1"
repeat_region	21809. .21914
	/rpt_family="L1"
repeat_region	22134. .22441
	/rpt_family="L1"
repeat_region	22575. .22911
	/rpt_family="L1"
repeat_region	23119. .23152
	/rpt_family="L1"
repeat_region	23403. .23660
	/rpt_family="L1"
repeat_region	23630. .24163
	/rpt_family="L1"
repeat_region	24506. .24935
	/rpt_family="L1"
repeat_region	25068. .25091
	/rpt_family="L1"
repeat_region	25684. .25801
	/rpt_family="L1"
repeat_region	26111. .26273
	/rpt_family="L1"
repeat_region	26396. .26932
	/rpt_family="L1"
repeat_region	27137. .27233
	/rpt_family="L1"
repeat_region	27944. .28009
	/rpt_family="L1"
repeat_region	28249. .28544
	/rpt_family="L1"
repeat_region	30553. .30690
	/rpt_family="L1"
repeat_region	30691. .30980
	/rpt_family="L1"
repeat_region	30981. .31085
	/rpt_family="L1"
repeat_region	31517. .31926
	/rpt_family="L1"
repeat_region	32248. .32599
	/rpt_family="L1"
repeat_region	32684. .32727
	/rpt_family="L1"
repeat_region	32807. .32882
	/rpt_family="L1"
repeat_region	33060. .33192
	/rpt_family="L1"
repeat_region	34040. .34137
	/rpt_family="L1"
repeat_region	34362. .34668
	/rpt_family="L1"
repeat_region	34768. .34788

```

misc_feature      /rpt_family="(TAAAA)n"
35045..35498
/notes="match to EST T03089 (NID:g314330)"
repeat_region    36105..36309
/rpt_family="L2"
repeat_region    37795..37817
/rpt_family="AT-rich"
repeat_region    38267..38459
/rpt_family="MaLR"
repeat_region    38470..38996
/rpt_family="MER21_g"
repeat_region    39037..39222
/rpt_family="MaLR"
repeat_region    39727..40330
/rpt_family="MER4-group"
repeat_region    41356..41531
/rpt_family="L2"
repeat_region    41563..41764

Query Match      90.6% Score 15.4; DB 11; Length 119483;
Best Local Similarity 94.1%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTAA 17
Db 51566 TCCTGGTCTGCAGGTCA 51550

RESULT 9
HSU55590/c      33414 bp DNA PRI 22-DEC-1997
LOCUS Homo sapiens IL-1 receptor antagonist IL-1ra (IL-1RN) gene,
DEFINITION alternatively spliced forms, complete cds.
ACCESSION U65590
VERSION g2707374
KEYWORDS U65590.1 GI:2707374
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 33414)
AUTHORS Jenkins,J.K., Drong,R.F., Shuck,M.E., Bienkowski,M.J.,
Slightom,J.L., Arend,W.P. and Smith,M.F. Jr.
TITLE Intracellular IL-1 receptor antagonist promoter: cell type-specific
and inducible regulatory regions
J Immunol. 158 (2), 748-755 (1997)
JOURNAL 97145044
MEDLINE 2 (bases 1 to 33414)
REFERENCE Slightom,J.L.
AUTHORS Direct Submission
TITLE Submitted (30-JUL-1996) Molecular Biology, Pharmacica & Upjohn
JOURNAL Company, 301 Hennretti, Kalamazoo, MI 49007, USA
FEATURES
source 1..33414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ic-1: P2g"
/chromosome="2"
/map="2q13-q14.1"
/dev_stage="adult"
1..236
/rpt_family="Alu"
/rpt_type="dispersed"
5614..5910
/rpt_family="Alu"
repeat_region /rpt_type="dispersed"
7319..7590
/rpt_family="Alu"
repeat_region /rpt_type="dispersed"
16075..16210,25876..25927,27756..27844,29053..29165,
30664..32022
/notes="IL-1RN"
mRNA

```

```
/codon_start=1
/product="IL-1 receptor antagonist IL-1Ra, secreted form"
/protein_id="AB92270.1"
/db_xref="PID:32707377"
/db_xref="GI:2707377"
/translation="MEICRGLRSLITLLFLFHSETICRPSGRKSKMQAFRIWDVN
QKTFYLRNNQLVAGYLOGPNVNEEKIDVPIEPHALFLGHHGKMLCSCKVSGDTR
LOLEAVNITDLSNRKQKRFAFIRSDSGPTTTFESAACPGWFCLCTAMEADQPVSLTN
MDEGVMTKFVQEDE"
25928..27755
/gene="IL-1RN"
/note="intron 1s"
27756..27844
/gene="IL-1RN"
/note="similar to GenBank Accession Numbers X52015 and
X53296"
/number=2
27845..29052
/gene="IL-1RN"
/number=2
28708..28880
/rpt_type=direct
/rpt_unit=28708..29794
29053..29165
/gene="IL-1RN"
/note="similar to GenBank Accession Numbers X52015 and
X53296"
/number=3
29166..30663
/gene="IL-1RN"
/number=3
30175..30462
/rpt_family="Alu"
/rpt_type=dispersed
30664..32022
/gene="IL-1RN"
/note="similar to GenBank Accession Numbers X52015 and
X53296"
/number=4
BASE COUNT 9153 a 7254 c 7854 g 9153 t
ORIGIN
|||||
1 TCCTGGTCTGCAGTAA 17
28914 TCCTGGTCTGCAGTAA 28898

Query Match 90.6%; Score 15.4; DB 11; Length 33414;
Best Local Similarity 94.1%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGTAA 17
|||||
Db 28914 TCCTGGTCTGCAGTAA 28898

RESULT 10
CNS00000T 159137 bp DNA HTG 14-MAY-1999
LOCUS Homo sapiens chromosome 14 clone bac R-112J1, WORKING DRAFT
DEFINITION SEQUENCE, in ordered pieces.
ACCESSION AL049873
VERSION AL049873.1 GI:4837625
KEYWORDS HTG; HTGS_PHASE2.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 159137)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1999) Genoscope - Centre national de sequence
2, rue Gaston Creneau - BP 191 91006 EVRY cedex - FRANCE (E-mail :
secref@genoscope.cns.fr)
COMMENT IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
```

```
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage, etc.
contigs_order:18a_17_18b; 1000 N's separate segments Contig18a :
length 5987 bp
Contig17 : length 13090 bp
Contig18b : length 138060 bp.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Location/Qualifiers
1..159137
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="14"
/clone_lib="R-112J1"
/clone_lib="R-112J1"
/clone_lib="R-112J1"
BASE COUNT 45907 a 31091 c 31465 g 48651 t 2023 others
ORIGIN
|||||
1 TCCTGGTCTGCAGTAA 17
37994 TCCTGGTCTGCAGTAA 38010

Query Match 90.6%; Score 15.4; DB 34; Length 159137;
Best Local Similarity 94.1%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGTAA 17
|||||
Db 37994 TCCTGGTCTGCAGTAA 38010

RESULT 11
HSJ1069C8/c
LOCUS Homo sapiens chromosome 20 clone J1069C8, WORKING DRAFT SEQUENCE,
DEFINITION in unordered pieces.
ACCESSION AL078623
VERSION AL078623.1 GI:5051362
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 181832)
AUTHORS Burton, J.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquires:
humquery@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known: 800 n's separate
segments. Unfinished: dJ1069C8 Contig_ID: 00859 acc= Length:
13801 bp Unfinished: dJ1069C8 Contig_ID: 02162 acc= Length:
23805 bp Unfinished: dJ1069C8 Contig_ID: 01473 acc= Length: 21242 bp
Unfinished: dJ1069C8 Contig_ID: 01277 acc= Length: 8213 bp
Unfinished: dJ1069C8 Contig_ID: 02169 acc= Length: 12911 bp
Unfinished: dJ1069C8 Contig_ID: 00785 acc= Length: 3964 bp
Unfinished: dJ1069C8 Contig_ID: 02011 acc= Length: 30508 bp
Unfinished: dJ1069C8 Contig_ID: 02131 acc= Length: 9098 bp
Unfinished: dJ1069C8 Contig_ID: 00235 acc= Length: 5405 bp
Unfinished: dJ1069C8 Contig_ID: 00434 acc= Length: 8806 bp
Unfinished: dJ1069C8 Contig_ID: 00554 acc= Length: 3424 bp
Unfinished: dJ1069C8 Contig_ID: 00521 acc= Length: 5317 bp
Unfinished: dJ1069C8 Contig_ID: 00048 acc= Length: 1064 bp
Unfinished: dJ1069C8 Contig_ID: 00481 acc= Length: 6646 bp
Unfinished: dJ1069C8 Contig_ID: 01292 acc= Length: 3448 bp
```

Unfinished: dJ1069C8 Contig_ID: 01374 acc= Length: 3829 bp
Unfinished: dJ1069C8 Contig_ID: 01572 acc= Length: 1331 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
 Location/Qualifiers
 1..181832
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="J1069C8"

BASE COUNT 51194 a 33312 c 33415 g 50303 t 13608 others
ORIGIN

Query Match 90.6%; Score 15.4; DB 34; Length 181832;
Best Local Similarity 94.1%; Pred. No. 70;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCTGTGCTGCAGGTAA 17
||||||| |||||
Db 103578 TCTGTGCTACAGGTAA 103562

RESULT 12
AC007491/c 91047 bp DNA HTG 06-MAY-1999
LOCUS Homo sapiens chromosome 16 clone 182c10, WORKING DRAFT SEQUENCE, 78
DEFINITION unordered pieces.
ACCESSION AC007491
NID 94755175
VERSION AC007491.1 GI:4755175
KEYWORDS HTG; HTGS-PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 91047)
AUTHORS Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,
Han, C. and Deaven, L.
TITLE Sequencing of Human Chromosome 16ql2
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 91047)
AUTHORS Ricke, D.O.
TITLE Large Scale Sequence Analysis and Annotation with the Sequence
Comparison Analysis (SCAN) System
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 91047)
AUTHORS Bruce, D., Mundt, M., Doggett, N., Munk, C., Saunders, E., Robinson, D.,
Jones, M., Buckingham, J., Chasteen, L., Thompson, S., Goodwin, L.,
Bryant, J., Tesmer, J., Meincke, L., Longmire, J., White, S., Tatum, O.,
Campbell, C., Fawcett, J., Maltbie, M., Bussod, M., Sutherland, R.,
McMurry, K., Han, C. and Deaven, L.
TITLE Direct Submission
JOURNAL Submitted (06-MAY-1999) Center for Human Genome Studies, DOE Joint
Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 78 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
* 1 631: contig of 631 bp in length
* 532 1265: contig of 634 bp in length
* gap of unknown length

1266 1864: contig of 599 bp in length
1865 gap of unknown length
2528 3147: contig of 663 bp in length
gap of unknown length
3148 4188: contig of 620 bp in length
gap of unknown length
4189 5476: contig of 1041 bp in length
gap of unknown length
5477 6354: contig of 1288 bp in length
gap of unknown length
6355 7483: contig of 878 bp in length
gap of unknown length
7484 7598: contig of 1129 bp in length
gap of unknown length
7599 8520: contig of 115 bp in length
gap of unknown length
8520 9309: contig of 921 bp in length
gap of unknown length
9310 10154: contig of 790 bp in length
gap of unknown length
10155 11418: contig of 845 bp in length
gap of unknown length
10726 11980: contig of 571 bp in length
gap of unknown length
11419 12829: contig of 693 bp in length
gap of unknown length
11981 14109: contig of 562 bp in length
gap of unknown length
12830 15277: contig of 849 bp in length
gap of unknown length
14110 16263: contig of 1280 bp in length
gap of unknown length
15278 17231: contig of 1168 bp in length
gap of unknown length
16264 17335: contig of 986 bp in length
gap of unknown length
17232 18301: contig of 104 bp in length
gap of unknown length
17336 19224: contig of 966 bp in length
gap of unknown length
18302 20340: contig of 923 bp in length
gap of unknown length
19225 21454: contig of 1116 bp in length
gap of unknown length
20341 22486: contig of 1114 bp in length
gap of unknown length
21455 23399: contig of 1032 bp in length
gap of unknown length
22487 24121: contig of 913 bp in length
gap of unknown length
23400 24987: contig of 722 bp in length
gap of unknown length
24122 26151: contig of 866 bp in length
gap of unknown length
24988 27357: contig of 1164 bp in length
gap of unknown length
26152 28440: contig of 1206 bp in length
gap of unknown length
27358 29550: contig of 1083 bp in length
gap of unknown length
28441 30365: contig of 1110 bp in length
gap of unknown length
29551 31025: contig of 815 bp in length
gap of unknown length
30366 32018: contig of 660 bp in length
gap of unknown length
31026 34031: contig of 993 bp in length
gap of unknown length
32019 34627: contig of 2013 bp in length
gap of unknown length
34032 contig of 596 bp in length

identity: 50.79; identified by sequence similarity;

Query Match 88.2%; Score 15; DB 2; Length 13143;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 CTGGTCTGCAGTAA 17
Db 11795 CTGGTCTGCAGTAA 11809

RESULT 14
LOCUS AB016081 21935 bp DNA VRT 16-MAR-1999
DEFINITION Oryzias latipes OIGC6 gene for guanylyl cyclase C, complete cds.
ACCESSION AB016081
NID 94521168
VERSION AB016081.2 GI:45211168
KEYWORDS guanylyl cyclase C; OIGC6.
SOURCE Oryzias latipes DNA.
ORGANISM Oryzias latipes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Adrianichthyoidei;
Adrianichthyidae; Oryziinae; Oryzias.
REFERENCE 1 (sites)
AUTHORS Mantoku, T., Muramatsu, R., Nakauchi, M., Yamagami, S., Kusakabe, T. and Suzuki, N.
TITLE Sequence analysis of cDNA and genomic DNA, and mRNA expression of the medaka fish homolog of mammalian guanylyl cyclase C
JOURNAL J. Biochem. 125, 476-486 (1999)
REFERENCE 2 (bases 1 to 21935)
AUTHORS Mantoku, T., Muramatsu, R., Nakauchi, M., Yamagami, S., Kusakabe, T. and Suzuki, N.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-1998) to the DBJ/EMBL/GenBank databases. Norio Suzuki, Hokkaido University, Division of Biological Sciences, Graduate School of Science, Kita 10, Nishi 8, Kita-ku, Sapporo, Hokkaido 060-0810, Japan (E-mail:norio-sesci.hokudai.ac.jp, Tel:81-11-706-4908, Fax:81-11-746-1512)
COMMENT On Mar 26, 1999 this sequence version replaced gi:3327368.
Sequence updated (10-MAR-1999).
Location/Qualifiers
1. .21935
/organism="Oryzias latipes"
/db_xref="taxon:8090"
6892. .21764
/gene="OIGC6"
Join(6892..7096,8516..8628,9327..9394,10225..10440,11155..11273,11499..11601,12054..12168,13323..13464,13594..13676,14180..14288,14544..14625,15278..15380,15453..15509,15591..15662,15760..15864,15956..16042,17711..17843,17909..18046,18115..18191,18272..18363,18444..18602,19876..20068,20540..20714,20786..20884,20964..21058,21132..21208,21554..21764)
/gene="OIGC6"
/codon_start=1
/product="guanylyl cyclase C"
/protein_id="BAA76279.1"
/db_xref="PID:d1040024"
/db_xref="PID:94521169"
/db_xref="GI:45211169"
/translation="MSTFNLMLSVLCAGFVSACVQDQGCQMDGTIVNVILLEDESP
WSLKVGGQILAEIKDAINAEGMFNLTVNFEFNTLYRQGCITISAGEAKL
NKLMTYGLGCAVLGPTCTYATFAIADVEKFNLSPTIISAGSFGSCDYAMNLRLL
PARKISDFINFEWEKFTIKKWTAYVYKQPNTECLWYIGALEADGRFLNYSR
TILRHGDUKDUKDSOENRSLNLCGSPDOLKEVKNISDAADNLDILFDLYND
VYITNTSMPEKNVNLVLTMDPTRTYITKPDLTGNDTNDYMAAYHDAVLVGVQMRD
IARINPAEMGMYNTYFRNVSGIGGYKLDSDYGRDVFNSVYITVDKYYKIL
FSFDENRTKMDPSPTFIWKALPDDKPGSELETODIIVVVLGVTVAVATLAFIF
YQNRKDRLRKRWSHNPDLISLLENNEHNIVSLKIEERKMKQIRALYDKKIV
LKLKSDGNFKAQRIELNALLHIDYSLTKFTGVKFDGFGVFGVGGGSLRYV
LNDKVSYPGCTFMDWEFKISVYDIAGMSYLHASDIIQVHGLKLTNCVVDNRVVKI

FEATURES
source

gene

CDS

TDFGCNAFLSRHDLTWTAPEHLRKEGTSQKGDVYSFAIQCEIVLRRTFTYFASLKR
SEKLSRVITSYFRPDNLNLETASKEAEVYMLIKSWEEDPEKRPDKFNKVENLGLKIIS
KIHNDONESYMDNMRRLOMYSKNLEHLVEERTALYKAERDRDLNFMLLPRPVVKS
LKESGAVEPELYDEVTIYFSDIVGFTTLCQYSTPMEVVDMLNDIYKGFDSIVDHDVY
KVETIGDAYWVAGSLPNRNGNHAVDICRMALDILEFMGTQFLRHVLGVIPWIRIGVH
SGPCAAGVGVKMPRYCLFGDVTNASTMESTGHPRIHVSEPTIQLIQTNCKFEYE
MRGETYLGKGTENTYMLTGESQDDYDLPTPTTENVQRLQQLAHMLACLERRSRG
SVRRKKHADQGNKDEDESGVSSQPEYLHLATVDNTLSITFL"

BASE COUNT 6207 a 4588 c 4625 g 6515 t
ORIGIN

Query Match 88.2%; Score 15; DB 4; Length 21935;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGT 15
Db 20527 TCCTGGTCTGCAGGT 20541

RESULT 15
LOCUS HS1044017/c 122325 bp DNA PRI 09-OCT-1998
DEFINITION Human DNA sequence from clone 1044017 on chromosome Xp11.3-11.4
Contains GSS and STS, complete sequence.
ACCESSION AL023875
NID G3449106
VERSION AL023875.1 GI:3449106
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 122325)
Direct Submission
Submitted (09-OCT-1998) E-mail enquiries: humquery@sanger.ac.uk
Clone requests: clonerequest@sanger.ac.uk
On Aug 21, 1998 this sequence version replaced gi:3288011.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
IMPORTANT: This sequence is not the entire insert of clone 1044017.
It may be shorter because we only sequence overlapping sections
once, or longer because we arrange for a small overlap between
neighbouring submissions.
The true left end of clone 50A13 (292545) is at 122226 in this
sequence. This sequence has been finished according to sequence map
criteria as follows. An attempt is made to resolve all sequencing
problems, such as compressions and repeats, but not necessarily
within known annotated human repeat sequence elements (e.g. Alu).
Where the sequence is ambiguous, there is an annotation using the
'unsure' feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
1044017 is from the library RPC15 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/ VECTOR: pcYPAC2.

FEATURES
source

1. .122325
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/clone="1044017"
/map="pil-3-11.4"
/clone_lib="RPC15"
repeat_region 382..2592

repeat_region /note="L1 repeat: matches 2060. .4298 of consensus"
2788. .2815
repeat_region /note="14 copies 2 mer ta 89% conserved"
complement(2894. .3421)
repeat_region /note="L1 repeat: matches 5191. .4664 of consensus"
3426. .3539
repeat_region /note="L1 repeat: matches 5214. .5330 of consensus"
3555. .3890
repeat_region /note="L1 repeat: matches 1. .348 of consensus"
4119. .4351
repeat_region /note="L1 repeat: matches 381. .586 of consensus"
4430. .8034
repeat_region /note="L1 repeat: matches 589. .4151 of consensus"
8030. .10187
repeat_region /note="L1 repeat: matches 3232. .5390 of consensus"
10039. .10924
repeat_region /note="L1PA9 repeat: matches 1. .908 of consensus"
10931. .11389
repeat_region /note="L1PA15 repeat: matches 75. .529 of consensus"
11390. .11692
repeat_region /note="AluY repeat: matches 1. .301 of consensus"
11708. .12073
repeat_region /note="L1PA7 repeat: matches 516. .890 of consensus"
12081. .14318
repeat_region /note="L1 repeat: matches 3060. .5390 of consensus"
14201. .15013
repeat_region /note="L1MD2 repeat: matches 30. .872 of consensus"
complement(16051. .16359)
repeat_region /note="MPR33 repeat: matches 323. .1 of consensus"
17313. .17398
repeat_region /note="MIR repeat: matches 125. .212 of consensus"
complement(17490. .18370)
repeat_region /note="L1PA15 repeat: matches 897. .1 of consensus"
complement(18221. .20110)
repeat_region /note="L1 repeat: matches 5390. .3513 of consensus"
20133. .20205
repeat_region /note="MLT26 repeat: matches 316. .388 of consensus"
20154. .20205
repeat_region /note="MLT2A repeat: matches 402. .453 of consensus"
complement(20243. .20391)
repeat_region /note="AluSg repeat: matches 146. .1 of consensus"
complement(20515. .21058)
repeat_region /note="L1PA15 repeat: matches 904. .356 of consensus"
21058. .21632
repeat_region /note="L1 repeat: matches 3898. .4474 of consensus"
24600. .24811
repeat_region /note="MLT1G repeat: matches 18. .233 of consensus"
24925. .25073
repeat_region /note="MLT1F repeat: matches 380. .529 of consensus"
complement(25572. .25735)
repeat_region /note="MIR repeat: matches 221. .41 of consensus"
27606. .28048
repeat_region /note="MSTB repeat: matches 1. .424 of consensus"
29230. .29542
repeat_region /note="L1MCI repeat: matches 757. .1079 of consensus"
complement(30245. .30469)
repeat_region /note="MIR repeat: matches 260. .35 of consensus"
30903. .31128
repeat_region /note="MIR repeat: matches 12. .254 of consensus"
31940. .32320
repeat_region /note="MSTA repeat: matches 1. .426 of consensus"
32748. .33119
repeat_region /note="MPR25 repeat: matches 1712. .2086 of consensus"
33216. .33503
repeat_region /note="L1 repeat: matches 1695. .1988 of consensus"
33779. .34487
repeat_region /note="L1M1 repeat: matches 204. .917 of consensus"
34896. .34941
repeat_region /note="23 copies 2 mer tg 96% conserved"
complement(34984. .35711)
repeat_region /note="L1PA5 repeat: matches 867. .101 of consensus"
complement(35712. .36011)
repeat_region /note="AluSg repeat: matches 300. .1 of consensus"

repeat_region complement(36013. .39594)
/note="L1 repeat: matches 5356. .1821 of consensus"
repeat_region complement(39599. .40328)
/note="L1 repeat: matches 1294. .631 of consensus"
40710. .>41173
misc_feature /note="match: GSS A0017146 clone 2302K13"
41281. .41417
repeat_region /note="L1MA2 repeat: matches 919. .1055 of consensus"
41427. .42116
repeat_region /note="L1 repeat: matches 2155. .2869 of consensus"
42137. .42434
repeat_region /note="AluX repeat: matches 1. .299 of consensus"
42458. .42891
repeat_region /note="L1 repeat: matches 2906. .3339 of consensus"
42909. .43377
repeat_region /note="L1 repeat: matches 3479. .3945 of consensus"
complement(43397. .43536)
repeat_region /note="L1MB5 repeat: matches 675. .539 of consensus"
43535. .43777
repeat_region /note="L1MB5 repeat: matches 673. .922 of consensus"
44532. .44707
repeat_region /note="MLT1G repeat: matches 1. .186 of consensus"
44600. .45088
repeat_region /note="MLT1F repeat: matches 69. .541 of consensus"
complement(45124. .45227)
repeat_region /note="MLT1G repeat: matches 488. .377 of consensus"
complement(45442. .45778)
misc_feature /note="match: GSS A0020067 clone 2303M7"
complement(45485. .45567)
repeat_region /note="MLT1F repeat: matches 184. .102 of consensus"
46053. .46102
repeat_region /note="25 copies 2 mer at 80% conserved"
complement(46718. .47172)
misc_feature /note="match: GSS A0004493 clone 2294C7; match: GSS
A0006878 clone 2294C10"
46787. .46898
repeat_region /note="MIR repeat: matches 152. .262 of consensus"
47199. .47458
repeat_region /note="L1PA15 repeat: matches 642. .904 of consensus"
complement(48589. .48663)
repeat_region /note="MIR repeat: matches 120. .45 of consensus"
complement(49037. .49199)
repeat_region /note="L1ME1 repeat: matches 661. .495 of consensus"
49903. .50335
repeat_region /note="MLT1C repeat: matches 28. .466 of consensus"
complement(50441. .50744)
repeat_region /note="AluJn repeat: matches 298. .1 of consensus"
50803. .50826
repeat_region /note="12 copies 2 mer ta 96% conserved"
50844. .51059
repeat_region /note="L1MA4A repeat: matches 834. .1047 of consensus"
51090. .51719
repeat_region /note="L1 repeat: matches 4760. .5390 of consensus"
51575. .52464
repeat_region /note="L1PA15 repeat: matches 1. .897 of consensus"
52842. .53043
repeat_region /note="L1PA13 repeat: matches 697. .899 of consensus"
53910. .53978
repeat_region /note="L1MA2 repeat: matches 971. .1044 of consensus"
complement(53981. .54055)
repeat_region /note="MSTA repeat: matches 424. .346 of consensus"
complement(54054. .54363)
repeat_region /note="MSTA repeat: matches 310. .1 of consensus"
54889. .55140
repeat_region /note="MIR repeat: matches 2. .261 of consensus"
complement(55862. .55914)
repeat_region /note="MLT1F repeat: matches 517. .464 of consensus"
57446. .57479
repeat_region /note="17 copies 2 mer ca 88% conserved"
57483. .57522
repeat_region /note="20 copies 2 mer ga 90% conserved"
58284. .58487
repeat_region /note="L1 repeat: matches 1761. .1981 of consensus"

```

repeat_region 58677..59040
               /note="THEIC repeat: matches 3. .371 of consensus"
misc_feature   <58820..59202
               /note="match: STS G13783"

Query Match    88.2%; Score 15; DB 9; Length 122325;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TCCTGGTCTGCAGGT 15
         |||||
Db      3873 TCCTGGTCTGCAGGT 3859

```

Search completed: September 18, 1999, 15:49:15
Job time: 1435 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Search Date: September 18, 1999, 05:27:21 ; Search time 213.04 Seconds
(without alignments)
19.965 Million cell updates/sec

Title: US-09-037-472-6
Perfect score: 17
Sequence: 1 TCCTGGTCTGCAGGTAA 17
Scoring table: IDENTITY_NUC
Searched: 311585 seqs, 125096042 residues
Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	1 T89948	Human interleukin-
2	17	100.0	17	1 T89410	Human IL-1RN gene
3	17	100.0	17	1 V32394	Interleukin-1RN pr
4	17	100.0	17	1 V60231	PCR primer used to
5	17	100.0	17	1 V62391	IL-1 receptor anta
6	17	100.0	12565	1 V62392	Human interleukin-
7	17	100.0	17	1 V16622	Interleukin 1 (441
8	15	88.2	110000	1 V1209_02	Continuation (3 of
9	14.4	84.7	6306	1 Q04100	Sequence encoding
10	14.4	84.7	1226	1 Q04735	Gene fragment enco
11	14.4	84.7	1483	1 Q74413	Spinach cysteine s
12	14.4	84.7	5107	1 T29600	Human EHOC-1 cDNA.
13	14.4	84.7	2951	1 T36390	Soluble chitinase
14	14	82.4	46899	1 Q54386	T. niveum Cyclospo
15	14	82.4	3125	1 T66902	Human EHOC-17 prot
16	14	82.4	5480	1 X13069	Enterococcus faeca
17	13.8	81.2	3690	1 Q30849	Type III procollag
18	13.8	81.2	235	1 Q89416	Human aspartoacyla
19	13.8	81.2	5460	1 T16508	Vector pAC3Al cont
20	13.8	81.2	111	1 T37197	J4-4 heavy chain v
21	13.8	81.2	114	1 T37202	Human prostate pro
22	13.8	81.2	768	1 T84941	Human prostatic mu
23	13.8	81.2	114	1 V12596	VH251/gammal mutan
24	13.8	81.2	111	1 V12591	VH251/gammal mutan
25	13.8	81.2	114	1 V55987	VH251/gammal mutan
26	13.8	81.2	111	1 V65982	Human kidney amino
27	13.8	81.2	16595	1 Q77982	cDNA corresp. to r
28	13.4	78.8	1382	1 Q77982	Human kidney amino
29	13.4	78.8	2471	1 Q55146	Pseudomonas aerugi
30	13.4	78.8	7617	1 V14354	Plasmid pGen-PGKgf
31	13.4	78.8	500	1 V69630	Human secreted pro
32	13.4	78.8	346	1 V75168	Staphylococcus aur
33	13.4	78.8	19196	1 X03987	HIV-1 NC infectiou
34	13.4	78.8	14512	1 X20529	Polynucleotide seq
35	13.4	78.8	2599	1 X33957	Human HCMV repress
36	13	76.5	1079	1 Q11478	Plasmid pUR3001
37	13	76.5	1079	1 N60175	D-amino-acid-oxida
38	13	76.5	448	1 Q59458	Human brain expres
39	13	76.5	52298	1 Q47357	L5 mycobacteriophage
40	13	76.5	5994	1 Q65674	Sequence encoding
41	13	76.5	10660	1 Q84793	Spinocerebellar at
42	13	76.5	2781	1 Q94954	Chicken oocyte rec
43	13	76.5	23	1 Q90371	Mouse tie-2 recept

ALIGNMENTS

RESULT 1

T89948
ID T89948 standard; DNA; 17 BP.
AC T89948; 1998 (first entry)
DT 05-MAR-1998
DE Human Interleukin-1 receptor antagonist intron 2 PCR primer 2.
KW Interleukin-1 receptor antagonist; IL-1ra; ulcerative colitis;
diagnosis; prognosis; inflammatory bowel disease; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9725445-A1.
PD 17-JUL-1997.
PF 08-JAN-1997; U00042.
PR 12-JAN-1996; US-587911.
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
PI Cominelli F, Pizarro T, Rotter JI, Yang H;
WPI; 97-372880/34.
PT Screening for ulcerative colitis in subjects of Jewish ancestry - by
detecting allele 2 of the VNTR (variable number of tandem repeats)
PT polymorphism at intron 2 of the IL-1 receptor antagonist gene
PS Claim 7; Page 17; 22pp; English.
CC PCR primers T89947 and T89948 are used to amplify intron 2 of the
human interleukin-1 receptor antagonist gene (IL-1ra) in a novel method
CC to screen for ulcerative colitis (UC) in a subject of Jewish ancestry.
CC There is an association between allele 2 of the variable number of tandem
CC repeats (VNTR) polymorphism at intron 2 of IL-1ra, an important
CC endogenous regulator of inflammation, and UC in humans of Jewish
CC ancestry. This method can be used for the diagnosis and prognosis of UC
CC in Jewish patients for UC and distinguishing UC from Crohn's disease (CD)
CC and other inflammatory disease of the bowel.
SQ Sequence 17 BP; 3 A; 4 C; 5 G; 5 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTAA 17
|||||
Db 1 TCCTGGTCTGCAGGTAA 17

RESULT 2

T89410
ID T89410 standard; cDNA; 17 BP.
AC T89410; 1998 (first entry)
DT 22-APR-1998
DE Human IL-1RN gene intron 2 PCR primer 2.
KW Osteoporosis; interleukin-1 receptor antagonist; IL-1RN; allele;
bone mineral density; post-menopause; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9738135-A1.
PD 16-OCT-1997.
PF 03-APR-1997; U05626.
PR 05-APR-1996; US-628282.
PA (MEDI-) MEDICAL SCI SYSTEMS INC.
PI Duff GW, Eastell R, Russell G;
WPI; 97-512741/47.
PT Detecting genetic predisposition for osteoporosis - by detecting
PT interleukin-1 receptor antagonist gene IL-1RN allele 2 in the
PT genomic DNA of a patient
PS Claim 2; Page 9; 21pp; English.
CC T89410 and T89411 are PCR primers used to amplify a region of the
interleukin-1 receptor antagonist gene, IL-1RN, intron 2 which contains

CC a variable number tandem repeat (VNTR) region that gives rise to five
 CC alleles. This product is used for predicting the risk of osteoporosis in
 CC a subject by determining the allelic and genetic polymorphism pattern for
 CC IL-1RN in genomic DNA. A pattern of at least one copy of the IL-1RN
 CC allele 2 indicates an increased susceptibility to osteoporosis. The
 CC methods can predict low bone mineral density (BMD) and the rate of bone
 CC density loss and thereby a susceptibility to osteoporosis. Individuals
 CC so identified can then be treated more aggressively to prevent or retard
 CC the occurrence of disease.
 SQ Sequence 17 BP; 3 A; 4 C; 5 G; 5 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTAA 17
 |||||
 Db 1 TCCTGGTCTGCAGGTAA 17

RESULT 3

V32394 ID V32394 standard; DNA; 17 BP.
 AC V32394.
 DT 11-SEP-1998 (first entry)
 DE Interleukin-1RN primer 2.
 KW Genetic polymorphism; PCR; primer; amplification; interleukin-1RN;
 KW sight threatening diabetic retinopathy; interleukin-1-alpha; IL-1RN;
 KW interleukin-1-beta; ss.
 OS Synthetic.
 OS Homo sapiens.
 PN WO9815653-A1.
 PD 16-APR-1998.
 PF 09-OCT-1997; G02790.
 PR 10-OCT-1996; GB-021129.
 PA (DUFFE/) DUFFE G.
 PA (RENN/) RENNIE I.
 PA (RICH/) RICHARDSON R.
 PI Duff G, Rennie I, Richardson R;
 DR WPI: 98-240835/21.
 PT Predicting increased risk of sight-threatening diabetic retinopathy
 PT - comprises identifying genetic polymorphism pattern for genes
 PT IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
 PT symptoms occur

PS Claim 2; Page 33; 41pp; English.
 CC Interleukin-1RN (IL-1RN) primers 2 and 1 (V32393) were used to amplify
 CC the IL-1RN gene region to identify polymorphism of the VNTR region
 CC at the IL-1RN intron 2 locus. The invention claims to provide a method
 CC for predicting the risk of sight threatening diabetic retinopathy. The
 CC method involves isolating DNA from a patient and determining the DNA
 CC polymorphism pattern of the genes that code for interleukin-1-alpha,
 CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
 CC identified is then compared with controls of known DNA polymorphism
 CC patterns thereby identifying patients carrying a genetic polymorphism
 CC associated with increased risk of sight threatening diabetic retinopathy.
 CC The method may be able to identify diabetic patients at risk before the
 CC clinically detectable disorders occur. Polymorphism pattern
 CC determination of IL genes involved PCR reactions using primers
 CC V32389-V32398. The method is also claimed to be useful in conjunction
 CC with identification of other genes associated with sight threatening
 CC diabetic retinopathy in genomic DNA and therefore, in identifying
 CC diabetic patients expressing multiple risk patterns.
 SQ Sequence 17 BP; 3 A; 4 C; 5 G; 5 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTAA 17
 |||||
 Db 1 TCCTGGTCTGCAGGTAA 17

RESULT 4
 V60231 ID V60231 standard; DNA; 17 BP.
 AC V60231.
 DT 25-NOV-1998 (first entry)
 DE PCR primer used to amplify interleukin-1 receptor antagonist (IL-1RN).
 KW Interleukin-1 receptor antagonist; IL-1RN; predisposition;
 KW coronary artery disease; screen; PCR primer; ss.
 OS Synthetic.
 PN WO9840517-A1.
 PD 17-SEP-1998.
 PF 09-MAR-1998; U04725.
 PR 10-MAR-1997; US-813456.
 PA (MEDI-) MEDICAL SCI SYSTEMS INC.
 PI Crossman DC, Duff GW, Francis SE;
 DR WPI: 98-520829/44.
 PT Detection of predisposition to coronary artery disease - by
 PT comparative measurement of levels of expression of alleles from the
 PT interleukin 1 locus
 PS Claim 6; Page 15; 22pp; English.

CC PCR primers V60230-31 were used to amplify alleles associated with
 CC the interleukin-1 receptor antagonist (IL-1RN). The specification
 CC describes a method for determination of a patient's predisposition
 CC to coronary artery disease. The method comprises comparing an
 CC allele with a second allele which is predictive of coronary artery
 CC disease, where similarity between the first and second alleles
 CC indicates a predisposition to coronary artery disease. The method is
 CC used to genotype an individual's interleukin (IL)-1 loci, the
 CC overexpression of which correlates with coronary artery disease. The
 CC method is used to screen a patients' predisposition to coronary
 CC artery disease.
 SQ Sequence 17 BP; 3 A; 4 C; 5 G; 5 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
 Best Local Similarity 100.0%; Pred. No. 3.4;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTAA 17
 |||||
 Db 1 TCCTGGTCTGCAGGTAA 17

RESULT 5

V62391 ID V62391 standard; DNA; 17 BP.
 AC V62391.
 DT 19-JAN-1999 (first entry)
 DE IL-1 receptor antagonist gene intron 2 PCR primer #2.
 KW Human; interleukin-1 receptor antagonist gene; IL-1; polymorphism;
 KW diagnosis; osteoporosis; PCR primer; ss.
 OS Synthetic.

OS Homo sapiens.
 PN WO9844150-A1.
 PD 08-OCT-1998.
 PF 27-MAR-1998; G00944.
 PR (GEMI-) GEMINI RES LTD.
 PA Keen RW, Spector TD;
 PI WPI: 98-557135/47.
 DR WPI: 98-557135/47.

PT Diagnosis of osteoporosis by determining genotype of interleukin-1
 PT receptor antagonist gene - useful for diagnosing patient
 PT pre-disposition or susceptibility to osteoporosis and for
 PT therapeutic intervention
 PS Claim 9; Page 10; 36pp; English.

CC A method has been developed for the diagnosis of osteoporosis comprising
 CC determining the genotype of an interleukin-1 (IL-1) receptor antagonist
 CC gene (IL-1RN). The present sequence represents a PCR primer adapted to
 CC amplify a portion of intron 2 of an IL-1RN for use in the method of the
 CC invention. The method can be used for the diagnosis of disease,
 CC including diagnosis of osteoporosis and predisposition or susceptibility

CC to osteoporosis and for therapy.
SQ Sequence 17 BP; 3 A; 4 C; 5 G; 5 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTAA 17
|||||
Db 1 TCCTGGTCTGCAGGTAA 17

RESULT 6
ID V62392/c
AC V62392 standard; DNA; 12565 BP.
DT 19-JAN-1999 (first entry)
DE Human interleukin-1 receptor antagonist gene.
KW Human; interleukin-1 receptor antagonist gene; IL-1; polymorphism;
KW diagnosis; osteoporosis; ds.
OS Homo sapiens.
PN W09844150-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; G00944.
PR 27-MAR-1997; GB-006359.
PA (GEMI-) GEMINI RES LTD.
PI Keen RW, Spector TD;
DR WPI: 98-557135/47.
PT Diagnosis of osteoporosis by determining genotype of interleukin-1
PT receptor antagonist gene - useful for diagnosing patient
PT pre-disposition or susceptibility to osteoporosis and for
PT therapeutic intervention
PS Disclosure: Page 21-27; 36pp; English.
CC A method has been developed for the diagnosis of osteoporosis comprising
CC determining the genotype of an interleukin-1 (IL-1) receptor antagonist
CC gene (IL-1RN). The present sequence represents the human interleukin-1
CC receptor antagonist gene. The method can be used for the diagnosis of
CC disease, including diagnosis of osteoporosis and predisposition or
CC susceptibility to osteoporosis and for therapy.
SQ Sequence 12565 BP; 3217 A; 2980 C; 3072 G; 3294 T;

Query Match 100.0%; Score 17; DB 1; Length 12565;
Best Local Similarity 100.0%; Pred. No. 4.2;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTAA 17
|||||
Db 9279 TCCTGGTCTGCAGGTAA 9263

RESULT 7
ID X16622 standard; DNA; 17 BP.
AC X16622;
DT 29-APR-1999 (first entry)
DE Interleukin 1 (44112332) haplotype PCR primer #16.
KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;
KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;
KW Graves disease; systemic lupus erythematosus; lichen sclerosis;
KW ulcerative colitis; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN W09854359-A1.
PD 03-DEC-1998.
PF 21-MAY-1998; G01481.
PR 29-MAY-1997; GB-011040.
PA (CAMP/) CAMP N J.
PA (COXA/) COX A.
PA (DGIO/) DE GIOVINE F S.
PA (DUFF/) DUFF G.
PA Camp NJ, Cox A, De Giovine FS, Duff G;

DR WPI: 99-080814/07.
PT New method of determining a patient's susceptibility to inflammatory
PT disorders - by detecting the presence of an IL-1 (44112332)
PT haplotype, useful in designing treatment strategies that modulate
PT the activity of proteins produced by the IL-1 gene cluster
PS Claim 3; Page 33; 49pp; English.
CC A method has been developed for determining a patient's susceptibility
CC to an inflammatory disorder. The method comprises the detection of an
CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the
CC patient, where its presence indicates susceptibility to an inflammatory
CC disorder. X16607 to X16631 represent PCR primer used in the method for
CC detecting the IL-1 (44112332) haplotype. The method provides kits for
CC the early prediction of a patient's susceptibility to inflammatory
CC disorders, including coronary artery disease, osteoporosis, nephropathy
CC in diabetes mellitus, alopecia areata, graves disease, systemic lupus
CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of
CC alleles of the haplotype can be applied to particular inflammatory
CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,
CC psoriasis, and insulin dependent diabetes. The identification of a
CC disease-associated haplotype enables determination of which alleles are
CC causative, and this information is useful in designing treatment
CC strategies, including gene therapy and treatment using various agents
CC that modulate the activity of proteins produced by the IL-1 gene cluster.
CC Some alleles from the IL-1 gene cluster are associated with particular
CC inflammatory diseases, and insufficient IL-1 production appears to act
CC centrally in the pathology of these diseases. Therefore, the use of IL-1
CC gene clusters is useful in determining genetic susceptibility to
CC inflammatory diseases, including those with a multifactorial etiology
CC with a polygenic component.
SQ Sequence 17 BP; 3 A; 4 C; 5 G; 5 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 3.4;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTAA 17
|||||
Db 1 TCCTGGTCTGCAGGTAA 17

RESULT 8
ID V21209_02/c
AC Continuation (3 of 17) of V21209 from base 200001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments LOCUS V21209 Accession V21209
WP Fragment Name Begin End
WP V21209_00 1 110000
WP V21209_01 100001 210000
WP V21209_02 200001 310000
WP V21209_03 300001 410000
WP V21209_04 400001 510000
WP V21209_05 500001 610000
WP V21209_06 600001 710000
WP V21209_07 700001 810000
WP V21209_08 800001 910000
WP V21209_09 900001 1010000
WP V21209_10 1000001 1110000
WP V21209_11 1100001 1210000
WP V21209_12 1200001 1310000
WP V21209_13 1300001 1410000
WP V21209_14 1400001 1510000
WP V21209_15 1500001 1610000
WP V21209_16 1600001 1664976

Query Match 88.2%; Score 15; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 47;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CTGGTCTGCAGGTAA 17
|||||
Db 98370 CTGGTCTGCAGGTAA 98356

RESULT 9
Q04100
ID Q04100 standard; DNA: 6306 BP.
AC Q04100;
DT 13-SEP-1990 (first entry)
DE Sequence encoding human alpha-2-plasmin inhibitor.
KW Alpha-2-plasmin; alpha-2-plasmin inhibitor deficiency disease; ds.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT exon 881..899
FT /tag= a
FT intron 900..1936
FT /tag= b
FT exon 1937..2003
FT /tag= c
FT intron 2004..2108
FT /tag= d
FT exon 2109..2147
FT /tag= e
FT intron 2148..2267
FT /tag= f
FT exon 2268..2330
FT /tag= g
FT intron 2331..2640
FT /tag= h
FT exon 2641..2842
FT /tag= i
FT intron 2843..3821
FT /tag= j
FT exon 3822..3965
FT /tag= k
FT intron 3966..4111
FT /tag= l
FT exon 4112..4315
FT /tag= m
FT intron 4316..4498
FT /tag= n
FT exon 4499..4641
FT /tag= o
FT intron 4642..4752
FT /tag= p
FT exon 4753..4957
FT /tag= q
FT intron 4754..5137
FT /tag= r
FT exon 5138..5547
FT /tag= s
J02086778-A.
PD 27-MAR-1990. 234930
PR 21-SEP-1988; JP-234930
PA (TEIJ) Teijin KK.
DR WPI: 90-137117/18.
DR P-PSDB: R04252.
PT Human alpha-2-plasmin inhibitor protein coding gene -
PT has specified exons bonded to each other via intron.
PS Claim 1; Page 451; 10pp; Japanese.
CC Protein is derived from 10 exons of the gene widely spaced by
CC introns up to 6.0Kb in length. When inserted into a suitable
CC expression vector host animal cells can be used to produce human
CC alpha-2-plasmin inhibitor, useful in diagnosis and therapy of the
CC plasmin inhibitor deficiency disease.
SQ Sequence 6306 BP; 1146 A; 1861 C; 1961 G; 1338 T;

Query Match 84.7%; Score 14.4; DB 1; Length 6306;
Best Local Similarity 93.8%; Pred. No. 89;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 CCTGGTCTGCAGGTAA 17
|||||

Db 246 CCTGGTCTGCAGGTCA 261
RESULT 10
Q04735
ID Q04735 standard; DNA: 1226 BP.
AC Q04735;
DT 12-OCT-1990 (first entry)
DE Gene fragment encoding human alpha-2 plasmin inhibitor.
KW Alpha-2 plasmin; alpha-2-Pi; ds.
OS Homo sapiens.
PN J02119782-A.
PD 7-MAY-1990.
PR 31-OCT-1988; 272903
PR 31-OCT-1988; JP-272903.
PA (TEIJ) Teijin KK.
DR WPI: 90-182383/24.
PT Gene fragment -
PT is useful for expression of human alpha-2 plasmin inhibitor, etc.
PS Disclosure; P; Japanese.
SQ Sequence 1226 BP; 188 A; 360 C; 418 G; 260 T;
Query Match 84.7%; Score 14.4; DB 1; Length 1226;
Best Local Similarity 93.8%; Pred. No. 84;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 CCTGGTCTGCAGGTAA 17
|||||
Db 246 CCTGGTCTGCAGGTCA 261
RESULT 11
Q74413
ID Q74413 standard; cDNA to mRNA; 1483 BP.
AC Q74413;
DT 12-JUN-1995 (first entry)
DE Spinach cysteine synthase.
KW Cysteine; spinach; spinacia oleracea; pUC19; M13mpl8; plant; nutrition;
KW feed value; pharmaceutical; food additive; cosmetic; E.coli; ds.
OS Spinacia oleracea.
FH Key
FT Location/Qualifiers
FT cds 62..1213
FT /tag= a
FT /product= cysteine synthetase
J06245773-A.
PD 06-SEP-1994.
PR 26-FEB-1993; 038527.
PR 26-FEB-1993; JP-038527.
PA (MITS) MITSUBISHI CORP.
PA (MITU) MITSUBISHI KASEI CORP.
DR WPI: 94-321282/40.
DR P-PSDB: R63756.
PT Gene encoding a cysteine synthesis enzyme - useful for elevating
PT cysteine content in plant bodies increasing their nutritional
PT valve
PS Claim 2; Page 4-6; 6pp; Japanese.
CC The nucleotide sequence of the novel cysteine synthase enzyme from
CC Spinach, Spinacia oleracea. The probes (Q74414-5) were used to obtain a
CC clone from a cDNA library derived from spinach seedling leaves. The
CC 1.5 kb insert was ligated into the cloning vectors pUC19 and M13mpl8.
CC The gene, 1483 bp, encodes a protein of 383 a.a. The cysteine gene is
CC expressed in plants to elevate the cysteine content in the plant. The
CC nutritional and feed value of the plant are expected to be enriched.
CC Cysteine, one of the S-containing amino acids, can be used as materials
CC for various pharmaceuticals, food additives or cosmetics. Production of
CC the protein is a useful step in the production of cysteine.
SQ Sequence 1483 BP; 440 A; 272 C; 354 G; 417 T;
Query Match 84.7%; Score 14.4; DB 1; Length 1483;
Best Local Similarity 93.8%; Pred. No. 85;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY      2 CCTGGTCTGCAGGTAA 17
      ||||| |||||
Db      839 CCTGGTCTGCAGGTAA 854

RESULT 12
T29600/c
ID      T29600 standard; cDNA; 5107 BP.
AC      T29600;
DT      12-AUG-1996 (first entry)
DE      Human EHOC-1 cDNA.
KW      EHOC-1; chromosome 21; gene marker; genome; mutation; aneuploidy;
KW      progressive myoclonus epilepsy; PME; homoprosencephaly; HPE1;
KW      autoimmune polyglandular disease type I; APECED; artisenase;
KW      gene therapy; diagnosis; transgenic animal; ds.
OS      Homo sapiens.
FH      Key      Location/Qualifiers
FT      cds      138..3710
FT      /*tag= a
FT      /*label= EHOC-1
PN      W09615144-A2.
PD      23-MAY-1996.
PF      08-NOV-1995; U14641.
PR      09-NOV-1994; US-337690.
PA      (CEDA-) CEDARS SINAI MEDICAL CENT.
PI      Korenberg JR, Yamakawa K;
DR      WPI: 96-259777/26.
DR      P-PSDB: R95241.
PT      Human EHOC-1 gene, chromosome 21 gene marker - useful to prepare
PT      probes for detection of mutation(s) and aneuploidies in chromosome
PT      21 locus q22.3
PS      Claim 5; Page 39-45; 62pp; English.
CC      The human EHOC-1 gene (T29600) is derived from the q22.3 locus of
CC      chromosome 21, which is the site of mutations that cause progressive
CC      myoclonus epilepsy of the Unverricht Lundborg type (EPM1), autoimmune
CC      polyglandular disease type I (APECED) and holoprosencephaly (HPE1).
CC      It was obtd. by construction of a bacterial artificial chromosome
CC      contig of a EPM1-APECED-HPE1 candidate region and use of a direct
CC      cDNA selection technique. The gene can be used for prodn. of
CC      EHOC-1 polypeptide (R95241), to design diagnostic probes and primers,
CC      to supply wild-type gene function to a mutated EHOC-1 gene, to design
CC      antisense sequences for therapeutic appln., and to breed transgenic
CC      animal models of disease.
SQ      Sequence 5107 BP; 1237 A; 1322 C; 1324 G; 1224 T;

      Query Match      84.7%; Score 14.4; DB 1; Length 5107;
      Best Local Similarity 93.8%; Pred. No. 88;
      Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCTGGTCTGCAGGTA 16
      | ||||| |||||
Db      1764 TGCTGGTCTGCAGGTA 1749

RESULT 13
T36390/c
ID      T36390 standard; DNA; 2951 BP.
AC      T36390;
DT      14-JAN-1997 (first entry)
DE      Soluble chitinase coding sequence (chiA gene).
KW      Periplasmic chitodextrinase; periplasmic Beta-N-acetylglucosaminidase;
KW      Beta-N-acetylglucosaminidase; chitin; oligosaccharide; catabolic;
KW      catabolism; ss.
OS      Vibrio furnissii.
FH      Key      Location/Qualifiers
FT      cds      77..2677
FT      /*tag= a
FT      /*product= Soluble chitinase
PN      W09625424-A1.
PD      22-AUG-1996.
PF      13-FEB-1996; U02332.

      Query Match      84.7%; Score 14.4; DB 1; Length 5107;
      Best Local Similarity 93.8%; Pred. No. 88;
      Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCTGGTCTGCAGGTA 16
      ||||| ||||| |||||
Db      1764 TGCTGGTCTGCAGGTA 1749

      Query Match      84.7%; Score 14.4; DB 1; Length 2951;
      Best Local Similarity 93.8%; Pred. No. 87;
      Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TCCTGGTCTGCAGGTA 16
      ||||| ||||| |||||
Db      989 TCCAGGTCTGCAGGTA 974

RESULT 14
Q54386
ID      Q54386 standard; DNA; 46899 BP.
AC      Q54386;
DT      08-JUL-1994 (first entry)
DE      T. niveum Cyclosporin synthetase gene.
KW      Enzyme; cyclosporin; synthetase-like activity; Tolypocladium niveum;
KW      T. inflatum GAMS; biosynthesis; vector; cyclosporin synthetase; ss.
OS      Tolypocladium niveum.
FH      Key      Location/Qualifiers
FT      cds      885..46730
FT      /*tag= a
FT      /*product= Cyclosporin synthetase
FT      misc_feature 40239..43129
FT      /*tag= b
FT      /*note= "SalI restriction fragment, preferred
FT      fragment, Claim 4"
FT      37781..40244
FT      /*tag= d
FT      /*note= "SalI restriction fragment, preferred
FT      fragment, Claim 5"
EP      EP-578616-A.
PN      12-JAN-1994.
PD      05-JUL-1993; 810474.
PF      09-JUL-1992; AT-001403.
PR      08-MAR-1993; AT-000437.
PR      29-APR-1993; CH-001310.
PR      04-MAY-1993; CH-001375.
PA      (SANO ) SANDOZ LTD.
PA      (SANO ) SANDOZ PATENT GMBH.
PI      Leitner E, Schneider E, Schoergendorfer K, Weber G;
DR      WPI: 94-010432/02.
DR      P-PSDB: R44928.
PT      Isolated DNA sequence - which codes for enzyme having cyclosporin
PT      synthetase like activity
PS      Claim 6; Page 17-41; 93pp; English.
CC      This sequence encodes an enzyme which has cyclosporin synthetase-
CC      like activity. This sequence was isolated from Tolypocladium niveum
CC      (formerly known as T. inflatum GAMS). The enzyme encoded by this
CC      sequence catalyses the peptide biosynthesis of cyclosporins and
CC      structurally related molecules. This sequence may be used for the
CC      production of cyclosporin by transforming a vector containing this
CC      sequence in to a recombinant host. This allows effective production

```

CC of antibiotic cyclosporin or its derivatives.
SQ Sequence 46899 BP; 10651 A; 13513 C; 12509 G; 10226 T;

Query Match 82.4%; Score 14; DB 1; Length 46899;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGG 14
|||||
Db 20672 TCCTGGTCTGCAGG 20685

RESULT 15

T66902
ID T66902 standard: cDNA; 3125 BP.
AC T66902;
DT 20-JUL-1997 (first entry)
DE Human EHOC-17 protein cDNA clone.
KW EHOC-17; progressive myoclonus epilepsy; holoprosencephaly;
KW autoimmune polyglandular disease type I; transgenic animal;
KW chromosome 21; ion channel; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 25..2784
FT /*tag= a
PN WO9717437-A2.
PD 15-MAY-1997.
PF 08-NOV-1996; U17989.
PR 10-NOV-1995; US-006453.
PR 07-NOV-1996; US-006453.
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PI Korenberg JR, Yamakawa K;
DR WPI; 97-281026/25.
DR P-PSDB; W15197.
PT Nucleic acid encoding EHOC-17 peptide - located on chromosome 21 in
PT the region implicated in progressive myoclonus epilepsy and other
PT inherited disorders
PS Claim 3: Page 38-40; 50pp; English.
CC A cDNA clone (T66902) codes for novel human EHOC-17 protein
CC (W15197), an acetylcholine-activated cationic channel permeable to
CC Ca2+ ions. A cDNA library from a 14-wk trisomy 21 foetal brain was
CC screened with bacterial artificial chromosomes covering the 21q22.3
CC region. Specifically bound material was amplified by PCR and used
CC to probe the cDNA library, yielding the EHOC-17 clone. The clone
CC maps proximal to D21S25, i.e. within the consensus region for the
CC genes implicated in progressive myoclonal epilepsy, autoimmune
CC polyglandular disease type I and holoprosencephaly. A transcript
CC of the EHOC-17 gene is expressed in a wide range of adult tissues.
CC EHOC-17 oligonucleotides can be used to detect mutations and
CC aneuploidies at 21q22.3, while antisense oligonucleotides can be
CC used to modulate expression of EHOC-17 polypeptide. Cells
CC expressing EHOC-17 can be used in screening assays to identify
CC potential therapeutic agents. Transgenic animals can be used to
CC study the physiological and behavioral functions of EHOC-17.
SQ Sequence 3125 BP; 688 A; 845 C; 928 G; 664 T;

Query Match 82.4%; Score 14; DB 1; Length 3125;
Best Local Similarity 100.0%; Pred. No. 1.4e-02;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 CCTGGTCTGCAGGT 15
|||||
Db 273 CCTGGTCTGCAGGT 286

Search completed: September 18, 1999, 05:27:23
Job time: 1688 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:40 : Search time 1405 Seconds
(without alignments)
23.867 Million cell updates/sec

Title: US-09-037-472-6

Perfect score: 17
Sequence: 1 TCGTGGCTGCAGGTAA 17

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	15.4	90.6	490	44	AI286716
2	15.4	90.6	695	48	AI585162
3	15.4	90.6	406	48	AI616774
C 4	15	88.2	419	27	AA009390
C 5	15	88.2	593	28	C18373
C 6	15	88.2	379	29	AA193570
C 7	14.4	84.7	468	22	H09124
C 8	14.4	84.7	445	22	R59129
C 9	14.4	84.7	498	22	R59693
10	14.4	84.7	917	26	W26640
11	14.4	84.7	576	27	AA038299
12	14.4	84.7	634	29	AA125309
13	14.4	84.7	583	29	AA155819
C 14	14.4	84.7	288	30	AA227533
15	14.4	84.7	407	31	AA303078
C 16	14.4	84.7	438	31	AA311957
C 17	14.4	84.7	414	33	AA420782
18	14.4	84.7	484	33	AA429235
C 19	14.4	84.7	479	33	AA449817
C 20	14.4	84.7	446	33	AA451654
C 21	14.4	84.7	405	34	AA457189
C 22	14.4	84.7	340	34	AA470498
C 23	14.4	84.7	603	35	AA555838
C 24	14.4	84.7	336	37	AA666248
25	14.4	84.7	693	38	AA780625
26	14.4	84.7	618	38	AA800216
27	14.4	84.7	359	39	AA849593
C 28	14.4	84.7	390	39	AA866085
C 29	14.4	84.7	435	39	AA893029
C 30	14.4	84.7	665	41	AI054743
C 31	14.4	84.7	493	42	AI111961
C 32	14.4	84.7	442	42	AI112270
C 33	14.4	84.7	440	42	AI137253
C 34	14.4	84.7	553	45	AA858817
C 35	14.4	84.7	288	45	AI396106
36	14.4	84.7	267	45	AI396107
37	14.4	84.7	286	46	AI410748
38	14.4	84.7	682	48	AI560858
39	14.4	84.7	406	49	AU056855
40	14.4	84.7	234	51	AI710703
41	14.4	84.7	301	51	AI715581
42	14.4	84.7	362	51	AU075473
43	14.4	84.7	507	54	HS0010802
44	14	82.4	423	20	T16114
45	14	82.4	386	23	H52729

ALIGNMENTS

RESULT 1
AI286716
LOCUS ub34d04.r1 Soares 2MbWt Mus musculus cDNA clone IMAGE:1395175 5',
DEFINITION mRNA sequence.
ACCESSION AI286716
NID 93926469
VERSION AI286716.1 GI:3926469

24-NOV-1998

```

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 490)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.

TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Dec 5, 1997 this sequence version replaced gi:2662843.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:906891
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 426.
Location/Qualifiers
1..490
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="19"
/clone_lib="IMAGE:1395175"
/clone_lib="Soares 2NMT"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/notes="Vector: pT7T3P-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTTACCAATGAGTGGAGCGCGCGCTTTTGTGTGTGTGTGTGTGTGTGTGT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M.Fatima Bonaldo."
BASE COUNT 116 a 116 c 114 g 144 t
ORIGIN

Query Match 90.6%; Score 15.4; DB 44; Length 490;
Best Local Similarity 94.1%; Pred. NO. 1.2e-02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTAA 17
|||||
Db 184 TCCTGGTCTGCAGGTAA 200

RESULT 2
AI585162 695 bp mRNA EST 06-APR-1999
LOCUS fb95a08.yl zebrafish WashU MPIMG EST Danio rerio cDNA 5' similar to
DEFINITION gb:J03464 PROCOLLAGEN ALPHA 2(I) CHAIN PRECURSOR (HUMAN):contains
PTR5.t3 PTR5 repetitive element ;, mRNA sequence.
ACCESSION AI585162
NID 94571059
VERSION AI585162.1 GI:4571059
KEYWORDS EST.
SOURCE zebrafish.

```

```

ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes;
Cyprinoidae; Cyprinidae; Rasbora; Danio.
1 (bases 1 to 695)
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M.,
Eddy,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y.,
Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R.,
Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
Waterston,R. and Wilson,R.

TITLE WashU Zebrafish EST Project 1998
JOURNAL Unpublished (1998)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3186954.

Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Matthew Clark. cDNA Library Arrayed by:
Matthew Clark. DNA Sequencing by: Washington University Genome
Sequencing Center Clone Distribution: Genome Systems, St. Louis,
Missouri (web address: www.genomesystems.com) (email contact:
info@genomesystems.com) and Research Genetics, Huntsville, Alabama
(web address: www.resgen.com) (email contact: info@resgen.com) and
Ressourcenzentrum Primatendatenbank, Berlin, Germany (web address:
www.rzpd.de)
Seq primer: T3 ET from Amersham
High quality sequence stop: 479.
Location/Qualifiers
1..695
/organism="Danio rerio"
/db_xref="taxon:7955"
/map="21q"
/clone_lib="Zebrafish WashU MPIMG EST"
/sex="mixed"
/tissue_type="26 somite embryos, adult livers, shield
stage embryos"
/lab_host="XLI-blue MRF"
/notes="Vector: pSPORT1; Site_1: NotI; Site_2: SalI; 1st
strand cDNA was primed with a Not I - oligo(dT)15 primer
[5'pGACTAGTCTAGATCGAGCGCGCCCTTTTGTGTGTGTGTGTGTGTGTGTGT3'],
double-stranded cDNA was ligated to Sal I adaptors (BRL),
digested with Not I and cloned into the Not I and Sal I
sites of the pSPORT1 vector (BRL). Library was constructed
by Matthew Clark (Lehrach lab; ICRF, London and Max Planck
Institut fuer Molekulare Genetik, Berlin). cDNAs for EST
analysis were selected following oligonucleotide
hybridization fingerprinting of arrayed clones from
zebrafish late somitogenesis (26 ss), adult liver or
embryonic shield stage (5.6 h) libraries. Fingerprint
data were used to computationally cluster cDNAs, and a
single cDNA from each cluster was chosen for sequencing.
In some cases multiple members of the same cluster were
sequenced to assess clustering parameters or single clones
were sequenced additional times to assess quality
control."
BASE COUNT 153 a 200 c 197 g 141 t 4 others
ORIGIN

Query Match 90.6%; Score 15.4; DB 48; Length 695;
Best Local Similarity 94.1%; Pred. NO. 1.2e-02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTAA 17
|||||
Db 107 TCCTGGTCTGCAGGTAA 123

RESULT 3

```

```

FEATURES
source

```

A1616774
 LOCUS zehna0590.seq.F Zebrafish Embryonic Heart cDNA Library Danio rerio
 DEFINITION cDNA 5', mRNA sequence.
 ACCESSION A1616774
 NID 94625941
 VERSION A1616774.1 GI:4625941
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Cyprinidae; Rasbora; Danio.
 REFERENCE 1 (bases 1 to 406)
 AUTHORS Ton, C., Mably, J.D., Dempsey, A.A., Hwang, D.M., Fishman, M.C. and Liew, C.C.
 TITLE Identification and Characterization of Expressed Sequence Tags from an Embryonic Zebrafish Heart cDNA Library
 JOURNAL Unpublished (1999)
 COMMENT On Jun 5, 1998 this sequence version replaced gi:3188789.
 Contact: Liew CC
 Department of Laboratory Medicine and Pathobiology
 University of Toronto
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169785650
 Email: liewcc@utcc.utoronto.ca
 PCR Primers
 FORWARD: 5' GCCAGCTCGAATTAAACCTCCTCACTAAGGG 3'
 BACKWARD: 5' CCAGTGAATGTAATACGACTCACTATAGGCG 3'
 Seq primer: 5' GAAATTAACCTCCTCACTAAGGG 3'.
 Location/Qualifiers
 1..406
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone_lib="zebrafish Embryonic Heart cDNA Library"
 /dev_stage="embryonic day 3 post-fertilization"
 /lab_host="E.coli XLI-Blue mrf"
 /note="Organ: heart; Vector: Lambda ZAP Express; Site:1: EcoRI; Site:2: XhoI; mRNA was purified from embryonic zebrafish hearts (3 day post-fertilization). cDNA was synthesized using a XhoI-Oligo dT adaptor-primer. EcoRI adaptors were ligated, followed by digestion with XhoI, for directional cloning into pre-digested lambda ZAP Express vector."
 BASE COUNT 75 a 102 c 134 g 95 t
 ORIGIN
 Query Match 90.6%; Score 15.4; DB 48; Length 406;
 Best Local Similarity 94.1%; Pred. No. 1.1e+02;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Oy 1 TCCTGGTCTGCAGGTA 17
 |||||
 Db 225 TCCTGGTCTGCAGGGA 241
 RESULT 4
 AA009390/c
 LOCUS AA009390 419 bp mRNA
 DEFINITION TgESTz08d12.r1 TgME49 Tachyzoite cDNA Toxoplasma gondii cDNA clone
 tgz08d12.r1 5', mRNA sequence.
 ACCESSION AA009390
 NID 91470456
 VERSION AA009390.1 GI:1470456
 KEYWORDS EST.
 SOURCE Toxoplasma gondii.
 ORGANISM Toxoplasma gondii
 Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Sarcocystidae; Toxoplasma.
 REFERENCE 1 (bases 1 to 419)

AUTHORS Hehl, A., Manger, I., Marra, M., Sibley, L.D., Ajioke, J.A., Aslett, M.A., Dietrich, N., Dubuque, T., Hillier, L., Kucaba, T., Wan, K.L., Waterston, R.H., and Boothroyd, J.
 TITLE WashU-Merck-Stanford-NIH Toxoplasma EST project
 JOURNAL Unpublished (1996)
 COMMENT On Apr 14, 1993 this sequence version replaced gi:693311.
 Contact: Marra M
 WashU-Merck EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: toxowatson.wustl.edu
 Clones are available from Genome Systems (genome@MO.NET); the library is available from John Boothroyd (jboothr@popserver.stanford.edu)
 Seq primer: T3
 High quality sequence stop: 302.
 Location/Qualifiers
 1..419
 /organism="Toxoplasma gondii"
 /strain="ME49, clone PDS"
 /db_xref="taxon:5811"
 /map="1. 1"
 /clone="tgzz08d12.r1"
 /clone_lib="TgME49 Tachyzoite cDNA"
 /lab_host="XLI-Blue MRF"
 /note="Vector: Lambda ZAP; Site:1: EcoRI; Site:2: XhoI; Toxoplasma parasites were grown in human foreskin fibroblasts cultures in vitro. The library was constructed by A. Hehl and I. Manger, Stanford University. cDNAs were synthesized from polyA mRNAs by oligo d(T) priming and directionally cloned into the EcoRI and XhoI sites of the Lambda zap vector using the ZAP-cDNA synthesis kit (Statagene). Warning: the library contains a small percentage of human cDNAs derived from the human host cells."
 BASE COUNT 98 a 104 c 128 g 89 t
 ORIGIN
 Query Match 88.2%; Score 15; DB 27; Length 419;
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 TCCTGGTCTGCAGGT 15
 |||||
 Db 341 TCCTGGTCTGCAGGT 327
 RESULT 5
 C18373/c
 LOCUS C18373 593 bp mRNA
 DEFINITION C18373 Human placenta cDNA (TFujiwara) Homo sapiens cDNA clone
 GEN-561C11 5', mRNA sequence.
 ACCESSION C18373
 NID g1579975
 VERSION C18373.1 GI:1579975
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 593)
 AUTHORS Fujiwara, T., Hirano, H., Hishigaki, H., Horie, M., Kawai, A., Kuga, Y., Kyushiki, H., Nagata, M., Okuno, S., Ozaki, K., Shimizu, F., Shimada, Y., Shinomiya, H., Suzuki, M., Takaichi, A., Takeda, S., Watanabe, T., Maekawa, H., Nakamura, Y. and Takahashi, E.
 TITLE Otsuka cDNA project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1393845.

Contact: Tsutomu Fujiwara
Otsuka GEN Research Institute
Otsuka Pharmaceutical Co., Ltd
463-10 Kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
Tel: 0886-65-2888
Fax: 0886-37-1035.
Location/Qualifiers
1..593
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="19p12-p13.1"
/clone="GEN-561c11"
/clone_lib="Human placenta cDNA (TFujiwara)"
/tissue_type="placenta"
169 a 122 C 163 g 137 t 2 others

BASE COUNT
ORIGIN

Query Match 88.2%; Score 15; DB 28; Length 593;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGT 15
|||||
Db 509 TCCTGGTCTGCAGGT 495

RESULT 6
AA193570
LOCUS
DEFINITION
zr42a04.s1 Soares NHMPu.S1 Homo sapiens cDNA clone IMAGE:666030 3'
similar to contains element L1 repetitive element ;, mRNA sequence.
ACCESSION
AA193570
NID
G1782971
VERSION
AA193570.1 GI:1782971
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 379)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On Nov 29, 1993 this sequence version replaced gi:501856.

TITLE
JOURNAL
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 292.
Location/Qualifiers
1..379
/organism="Homo sapiens"
/db_xref="GDB:5428062"
/db_xref="taxon:9606"
/clone="IMAGE:666030"
/clone_lib="Soares NHMPu.S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH108"
/note="Organ: mixed (see below); Vector: pT7f3D-Pac
(pharmacia) with a modified polylinker; Site.1: Not
Site.2: Eco RI; Equal amounts of plasmid DNA from three

FEATURES
source

normalized libraries (melanocyte 2NbHM, pregnant uterus
NBHPU, and fetal heart NBHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT 113 a 71 c 78 g 114 t 3 others
ORIGIN

Query Match 88.2%; Score 15; DB 29; Length 379;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CCTGGTCTGCAGGTA 16
|||||
Db 314 CCTGGTCTGCAGGTA 328

RESULT 7
H09124
LOCUS
DEFINITION
Y197h05.s1 Soares infant brain lN1B Homo sapiens cDNA clone
IMAGE:46496 3', mRNA sequence.
ACCESSION
H09124
NID
9873946
VERSION
H09124.1 GI:873946
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 468)
Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
Trevasakis,E., Waterston,R., Williamson,A., Wohldmann,P. and
Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
On May 5, 1995 this sequence version replaced gi:797863.

TITLE
JOURNAL
COMMENT

CONTACT: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 362
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1594 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 362.
Location/Qualifiers
1..468
/organism="Homo sapiens"
/db_xref="GDB:419037"
/db_xref="taxon:9606"
/clone="IMAGE:46496"
/clone_lib="Soares infant brain lN1B"
/sex="female"
/dev_stage="73 days post natal"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: whole brain; Vector: Lfamid BA; Site.1: Not
I; Site.2: Hind III; 1st strand cDNA was primed with a Not
I - oligo(dT) primer [5',
AAGTGGAGAGATTGCGGCCGAGCAATTTTTTTTTTTTTTTT 3'];
double-stranded cDNA was ligated to Hind III adaptors

FEATURES
source

(Pharmacia), digested with Not I and directionally cloned into the Not I and Hind III sites of the LfaIid BA vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 160 a 59 c 103 g 139 t 7 others

Query Match 84.7%; Score 14.4; DB 22; Length 468;
Best Local Similarity 93.8%; Pred. No. 4e+02; 1; Indels 0; Gaps 0;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTA 16
||||| ||||||| |||||

Db 320 TCCTAGTCTGCAGGTA 305

RESULT 8

R59129/c 84.7%; Score 14.4; DB 22; Length 468;
LOCUS y996c09.r1 Soares infant brain lN1B Homo sapiens cDNA clone
DEFINITION IMAGE:41341 5', mRNA sequence.

ACCESSION R59129

NID 9829824

VERSION R59129.1 GI:829824

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

AUTHORS Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 445)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

On May 8, 1995 this sequence version replaced gi:800177.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert Size: 1709

High quality sequence stops: 365 Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 1709 Std Error: 0.00

Seq primer: M13RP1

High quality sequence stop: 365.

Location/Qualifiers

1. .445

/organism="Homo sapiens"

/db_xref="CD8:413882"

/db_xref="taxon:9606"

/clone="IMAGE:41341"

/clone_lib="Soares infant brain lN1B"

/sex="female"

/dev_stage="73 days post natal"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: whole brain; Vector: LfaIid BA; Site_1: Not

I; Site_2: Hind III; 1st strand cDNA was primed with a Not

I - oligo(dT) primer [5'

AACTGGAAGAATTTCGGCGCCGAGGAATTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Hind III adaptors

(Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the LfaIid BA vector.

Library went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 117 a 98 c 139 g 90 t 1 others

ORIGIN

Query Match 84.7%; Score 14.4; DB 22; Length 445;
Best Local Similarity 93.8%; Pred. No. 3.9e-02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTA 16

||||| ||||||| |||||

Db 384 TCCTGGTCCGCAGGTA 369

RESULT 9

R59693/c 84.7%; Score 14.4; DB 22; Length 445;
LOCUS yHila03.s1 Soares infant brain lN1B Homo sapiens cDNA clone
DEFINITION IMAGE:42773 3', mRNA sequence.

ACCESSION R59693

NID 9830388

VERSION R59693.1 GI:830388

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

REFERENCE Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 498)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 296

Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: SP6

High quality sequence stop: 296.

Location/Qualifiers

1. .498

/organism="Homo sapiens"

/db_xref="GDB:415314"

/db_xref="taxon:9606"

/map="12"

/clone="IMAGE:42773"

/clone_lib="Soares infant brain lN1B"

/sex="female"

/dev_stage="73 days post natal"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: whole brain; Vector: LfaIid BA; Site_1: Not

I; Site_2: Hind III; 1st strand cDNA was primed with a Not

I - oligo(dT) primer [5'

AACTGGAAGAATTTCGGCGCCGAGGAATTTTTTTTTTTT 3'];

double-stranded cDNA was ligated to Hind III adaptors

(Pharmacia), digested with Not I and directionally cloned

into the Not I and Hind III sites of the LfaIid BA vector.

Library went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 159 a 69 c 115 g 146 t 9 others

ORIGIN

Query Match 84.7%; Score 14.4; DB 22; Length 498;
Best Local Similarity 93.8%; Pred. No. 4e-02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```

^~ 1 TCCTGGTCTGCAGGTA 16
    ||||| |||||
Db 321 TCCTAGTCTGCAGGTA 306

RESULT 10
W26640 917 bp mRNA EST 08-MAY-1996
LOCUS 34b6 Human retina cDNA randomly primed sublibrary Homo sapiens
DEFINITION cDNA, mRNA sequence.
ACCESSION W26640
NID 91307483
VERSION W26640.1 GI:1307483
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 917)
AUTHORS Macke, J., Smallwood, P. and Nathans, J.
TITLE Adult Human Retina cDNA
JOURNAL Unpublished (1996)
COMMENT On May 18, 1995 this sequence version replaced gi:811365.

Contact: Dr. Jeremy Nathans
Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics
Johns Hopkins School of Medicine
725 North Wolfe Street, Baltimore, MD 21205
Tel: 410 955 4678
Fax: 410 614 0827
Email: jeremy.nathans@gmail.bs.jhu.edu
Clones from this library are NOT available.
PCR Primers
FORWARD: CTTTTCACCAAGTTCAGCTGCTTAAGT
BACKWARD: GAGGTGGCTATGAGTATTTCTTCAGGGTAA
Seq primer: GGGTAAAGACCAAGAAATT.
Location/Qualifiers
1..917
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="Human retina cDNA randomly primed sublibrary"
/sex="mixed (males and females)"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="E. coli strain K802"
/notes="Organ: eye; Vector: lambda gt10; Site_1: EcoRI;
Site_2: EcoRI; The library used for sequencing was a
sublibrary derived from a human retina cDNA library.
Inserts from retina cDNA library DNAs were isolated,
randomly primed, PCR amplified, size-selected, and cloned
into lambda gt10. Individual plaques were arrayed and
used as templates for PCR amplification, and these PCR
products were used for sequencing."
BASE COUNT 124 a 162 c 127 g 150 t 354 others
ORIGIN
Query Match 84.7%; Score 14.4; DB 26; Length 917;
Best Local Similarity 93.8%; Pred. No. 4.4e-02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTA 16
||||| |||||
Db 279 TCCTGGTCTGCAGGTA 294

RESULT 11
AA038299 576 bp mRNA EST 28-AUG-1996
LOCUS mi83d03 rl Soares mouse p3NMF19.5 Mus musculus cDNA clone
DEFINITION IMAGE:473189 5' similar to gb:J04694 Mus musculus alpha-1 type IV
collagen (MOUSE);, mRNA sequence.
AA038299
NID 91684541
VERSION AA125309.1 GI:1684541
KEYWORDS EST.

ACCESSION AA038299
NID 91513706
VERSION AA038299.1 GI:1513706
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
REFERENCE 1 (bases 1 to 576)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 21, 1992 this sequence version replaced gi:279401.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:283933
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 417.
Location/Qualifiers
1..576
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:473189"
/clone_lib="Soares mouse p3NMF19.5"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/notes="Vector: pT73B (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTGAGTGGAGCGGCCCATTTTTTTTTTTT 3'],
double-stranded cDNA was size selected, ligated to Eco RI
adapters (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of a modified pT73 vector
(Pharmacia). Library went through one round of
normalization to a Cot = 5. Library constructed by Bento
Soares and M. Fatima Bonaldo. RNA was kindly provided by
Dr. Minoru Ko (Wayne State University)."
BASE COUNT 125 a 163 c 163 g 125 t
ORIGIN
Query Match 84.7%; Score 14.4; DB 27; Length 576;
Best Local Similarity 93.8%; Pred. No. 4.1e-02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TCCTGGTCTGCAGGTA 16
||||| |||||
Db 89 TCCTGGTCTGCAGGTA 104

RESULT 12
AA125309 634 bp mRNA EST 18-FEB-1997
LOCUS mp75c02.rl Soares 2NbWT Mus musculus cDNA clone IMAGE:575042 5'
DEFINITION Similar to gb:M1315 PROCOLLAGEN ALPHA 1(IV) CHAIN PRECURSOR
(HUMAN); gb:J04694 Mus musculus alpha-1 type IV collagen (MOUSE);,
mRNA sequence.
AA125309
NID g1684541
VERSION AA125309.1 GI:1684541
KEYWORDS EST.

```

SOURCE
house mouse
ORGANISM
Mus musculus

REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 634)

AUTHORS
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.

TITLE
The WashU-HMI Mouse EST Project

JOURNAL
Unpublished (1996)

COMMENT
On Sep 12, 1996 this sequence version replaced gi:1282806.

Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:349690
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 308.

FEATURES
Location/Qualifiers
1..634

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:575042"
/clone_lib="Soares 2NBM"
/sex="male"
/tissue_type="Thymus"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker. Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCGAGTGGAGCGCGGTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through two
rounds of normalization, and was constructed by Bento
Soares and M. Fatima Bonaudo."

BASE COUNT 141 a 182 c 186 g 124 t 1 others
ORIGIN

Query Match 84.7%; Score 14.4; DB 29; Length 634;
Best Local Similarity 93.8%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTA 16
|||||
Db 42 TCCTGGTCTGCAGGTA 57

RESULT 13
AA155819/c
LOCUS AA155819 583 bp mRNA EST 11-DEC-1996
DEFINITION z047c02.r1 Stratagene endothelial cell 937223 Homo sapiens cDNA
clone IMAGE:590018 5', mRNA sequence.

ACCESSION AA155819
NID g1727496
VERSION AA155819.1 GI:1727496
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 583)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevaskis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On Sep 12, 1996 this sequence version replaced gi:1309494.

TITLE
JOURNAL
MEDLINE
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 396.

FEATURES
Location/Qualifiers
1..583

/organism="Homo sapiens"
/db_xref="GDB:4621404"
/db_xref="taxon:9606"
/clone="IMAGE:590018"
/clone_lib="Stratagene endothelial cell 937223"
/dev_stage="umbilical vein, 1 passage"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Umbilical vein endothelial cells, passaged once. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCGCAGG 3' -3' adaptor sequence: 5'
CTCAGATTTTTTTTTTTT 3'"

BASE COUNT 162 a 119 c 190 g 104 t 8 others
ORIGIN

Query Match 84.7%; Score 14.4; DB 29; Length 583;
Best Local Similarity 93.8%; Pred. No. 4.1e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTA 16
|||||
Db 340 TCCTGGTCTGCAGGTA 325

RESULT 14
AA227533/c
LOCUS AA227533 288 bp mRNA EST 24-FEB-1997
DEFINITION zr57c08.r1 Soares_NHMPu_S1 Homo sapiens cDNA clone IMAGE:667502
5', mRNA sequence.

ACCESSION AA227533
NID g1849159
VERSION AA227533.1 GI:1849159
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 288)
REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,
Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and
Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1392732.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewatson.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 287.

FEATURES

Location/Qualifiers
1. .288
/organism="Homo sapiens"
/db_xref="CDB:5561469"
/db_xref="taxon:9606"
/clone="IMAGE:667502"
/clone_lib="Soares.NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not I;
Site_2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NbHPU, and fetal heart NbHH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

source

BASE COUNT 68 a 71 c 97 g 52 t
ORIGIN

Query Match 84.7%; Score 14.4; DB 30; Length 288;
Best Local Similarity 93.8%; Pred.No. 3.6e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTA 16
||||||| |||||
Db 258 TCCTGGTCCGCGAGTA 243

RESULT 15
AA303078 407 bp mRNA EST 18-APR-1997
LOCUS EST12915 Uterus tumor I Homo sapiens cDNA 3' end similar to
DEFINITION hypothetical protein, transmembrane (GB:U19252), mRNA sequence.
ACCESSION AA303078
NID 91955408
VERSION AA303078.1 GI:19555408
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 407)
Adams,M.D., Soares,M.B., Kerlavage,A.R., Fields,C. and Venter,J.C.
AUTHORS Rapid cDNA sequencing (expressed sequence tags) from a
TITLE directionally cloned human infant brain cDNA library
JOURNAL Nature Genet. 4, 373-380 (1993)
MEDLINE 94004965
COMMENT On May 8, 1995 this sequence version replaced gi:801297.
Other_ESTs: EST12914 THC173606
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlav@tigr.org

For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene
Index (<http://www.tigr.org/tldb/hgi/hgi.html>)
Seq primer: M13-21.

FEATURES

Location/Qualifiers
1. .407
/organism="Homo sapiens"
/db_xref="ATCC (inhost):192778"
/db_xref="taxon:9606"
/clone_lib="uterus tumor I"
/dev_stage="adult"
/note="Organ: uterus; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI"

BASE COUNT 86 a 88 c 102 g 129 t 2 others
ORIGIN

Query Match 84.7%; Score 14.4; DB 31; Length 407;
Best Local Similarity 93.8%; Pred.No. 3.9e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TCCTGGTCTGCAGGTA 16
||||||| |||||
Db 135 TCCTGGTCTGCAGGTA 150

Search completed: September 18, 1999, 06:47:43
Job time: 3660 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 15:48:58 ; Search time 436.05 Seconds
(without alignments)
123.989 Million cell updates/sec

Title: US-09-037-472-5
Perfect score: 17
Sequence: 1 CTCACACACTCTCTAT 17

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*

- 1: gb_bal.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_st.*
- 14: gb_sts.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vl.*
- 18: em_fun.*
- 19: em_htg.*
- 20: em_hum1.*
- 21: em_hum2.*
- 22: em_in.*
- 23: em_om.*
- 24: em_or.*
- 25: em_ov.*
- 26: em_pat.*
- 27: em_ph.*
- 28: em_pl.*
- 29: em_ro.*
- 30: em_sts.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vl.*
- 34: gb_htg1.*
- 35: gb_htg2.*
- 36: gb_in1.*
- 37: gb_in2.*
- 38: em_bal.*
- 39: em_ba2.*
- 40: em_hum3.*
- 41: em_hum4.*
- 42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	17	100.0	1105	4	AF031392	AF031392 Poecilia

RESULT 1
AF031392
LOCUS AF031392 1105 bp DNA
DEFINITION Poecilia perugiae NADH dehydrogenase subunit 2 gene, mitochondrial
ACCESSION AF031392
NID 92653588
VERSION AF031392.1 GI:2653588
KEYWORDS
SOURCE Poecilia perugiae.
ORGANISM Mitochondrion Poecilia perugiae
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Acanthopterygii; Poeciliidae; Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae; Poecilia.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden.F., Placek.M., Rashed.M., Taphorn.D. and de Figueiredo,C.A.
TITLE Molecular phylogeny of a live-bearing fish genus Poecilia (Cyprinodontiformes: Poeciliidae) based on NADH dehydrogenase subunit 2 sequence variation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden.F., Placek.M. and Rashed.M.
TITLE Direct Submission

ALIGNMENTS

2	17	100.0	1105	4	AF031393	AF031393 Poecilia
3	17	100.0	1105	4	AF031396	AF031396 Poecilia
4	17	100.0	1105	4	AF084973	AF084973 Heterand
5	17	100.0	1105	5	I85578	I85578 Sequence 1
6	17	100.0	12565	10	HSIL1RECA	X64532 H.sapiens g
7	17	100.0	33414	11	HSU65590	U65590 Homo sapien
8	16	94.1	1105	4	AF031389	AF031389 Poecilia
c 9	15.4	90.6	12777	2	AE000669	AE000669 Aquifex a
c 10	15.4	90.6	11891	2	AE000763	AE000763 Aquifex a
11	15.4	90.6	3360	2	ECY86689	U86689 Erwinia chr
12	15.4	90.6	1105	4	AF031387	AF031387 Poecilia
13	15.4	90.6	1105	4	AF031388	AF031388 Poecilia
14	15.4	90.6	1105	4	AF031390	AF031390 Poecilia
15	15.4	90.6	1105	4	AF031394	AF031394 Poecilia
16	15.4	90.6	1105	4	AF031400	AF031400 Poecilia
17	15.4	90.6	759	4	AF031401	AF031401 Poecilia
18	15.4	90.6	1105	4	AF031402	AF031402 Poecilia
19	15.4	90.6	1144	4	AF080487	AF080487 Poecilia
20	15.4	90.6	1124	4	AF080488	AF080488 Poecilia
21	15.4	90.6	1124	4	AF080489	AF080489 Poecilia
22	15.4	90.6	1124	4	AF080490	AF080490 Poecilia
23	15.4	90.6	1144	4	AF080491	AF080491 Poecilia
24	15.4	90.6	101072	11	AC005752	AC005752 Homo sapi
25	15.4	90.6	22710	12	D85605	D85605 Mouse DNA f
c 26	15	88.2	1167	7	HA2SALB	X78101 H.annuus mR
27	15	88.2	38785	10	HS695020B	AL049853 Human DNA
28	15	88.2	166180	34	HS695020	AL032818 Homo sapi
c 29	14.4	84.7	1727	2	AH14748	U14748 Aeromonas h
30	14.4	84.7	1105	4	AF031386	AF031386 Xiphophor
31	14.4	84.7	1105	4	AF031395	AF031395 Poecilia
32	14.4	84.7	624	4	CHKM3EN	J08893 chicken ovo
33	14.4	84.7	391	4	CHKOVOM2	J08898 chicken ovo
34	14.4	84.7	45	4	GGVO2	V00441 Chicken ovo
35	14.4	84.7	2323	7	GMSSLWSRP	Z18359 Glycine max
c 36	14.4	84.7	115359	8	ATAC003028	AC003028 Arabidops
37	14.4	84.7	6335	9	D86982	D86982 Human mRNA
c 38	14.4	84.7	323000	9	HS229041	AJ229041 Homo sapi
c 39	14.4	84.7	147055	9	HS372K1	AL023360 Human DNA
c 40	14.4	84.7	190837	10	HS90L6	Z97353 Human DNA s
c 41	14.4	84.7	39960	10	HSU162C4	Z80903 Human DNA s
c 42	14.4	84.7	172983	11	AC002454	AC002454 Human BAC
c 43	14.4	84.7	190907	11	AC005261	AC005261 Homo sapi
44	14.4	84.7	23179	11	AC06154	AC06154 Homo sapi
c 45	14.4	84.7	179544	11	AF015720	AF015720 Homo sapi

```
JOURNAL Submitted (27-OCT-1997) Biological Sciences, Simon Fraser
University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
FEATURES
  source
    1..1105
      /organism="Poecilia perugiae"
      /mitochondrion
      /db_xref="taxon:69232"
      /note="Tribe Poeciliini"
      33..1079
        /codon_start=1
        /transl_table=2
        /product="NADH dehydrogenase subunit 2"
        /protein_id="AAB87618.1"
        /db_xref="PID:g2653599"
        /db_xref="GI:2653598"
        /translation="MAPPYISALIIISGLGTTMTFASTHWYLAWMGIEINTLAIPLM
        AONHPRATEATKYFFAQATASATLLFASISNAFLTGEWDILHTPTSTLITLAL
        AMKIGLAPLHSMPEVMQGLNLTGLILSTWQKLAPLYLIYOQPNNTFTLGLLS
        IIVGGWGFNOVLKRLAYSSIAHLGWMILILSFSPPLALLTIYLLMTFSLFSF
        MLTRATHINSLTWAKIPILTYTPLLILSLGGLPLPGFMKWLILQELTKQDLAP
        LATLAALSLSLSLYFYRLSYAMTSLMPPNPNPAGTLPWRLNPQHNTFLMALTTSTIC
        LLPVTPAIMALLPF"
BASE COUNT 309 a 386 c 125 g 264 t 1 others
ORIGIN

Query Match 100.0%; Score 17; DB 4; Length 1105;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCAACTCTCTAT 17
    |||||
Db 230 CTCAGCAACTCTCTAT 246

RESULT 2
AF031393
LOCUS AF031393 1105 bp DNA VRT 30-NOV-1997
DEFINITION Poecilia reticulata NADH dehydrogenase subunit 2 gene,
mitochondrial gene encoding mitochondrial protein, complete cds.
ACCESSION AF031393
NID g2653590
VERSION AF031393.1 GI:2653590
KEYWORDS
SOURCE
  ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
    Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
    Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
    Poecilia.
REFERENCE 1 (bases 1 to 1105)
  Breden,F., Ptacek,M., Rashed,M., Taphorn,D. and de Figueiredo,C.A.
  Molecular phylogeny of a live-bearing fish genus Poecilia
  (Cyprinodontiformes: Poeciliidae) based on NADH dehydrogenase
  subunit 2 sequence variation
  Unpublished
REFERENCE 2 (bases 1 to 1105)
  Breden,F., Ptacek,M. and Rahsed,M.
  Direct Submission
  Submitted (27-OCT-1997) Biological Sciences, Simon Fraser
  University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
  Location/Qualifiers
    1..1105
      /organism="Poecilia reticulata"
      /mitochondrion
      /db_xref="taxon:8081"
      /note="Tribe Poeciliini; Oropuche River, Oropuche
      Drainage, Trinidad"
      33..1079
        /codon_start=1
        /transl_table=2
        /product="NADH dehydrogenase subunit 2"
        /protein_id="AAB87618.1"
CDS
  source
    1..1105
      /organism="Poecilia perugiae"
      /mitochondrion
      /db_xref="taxon:69232"
      /note="Tribe Poeciliini"
      33..1079
        /codon_start=1
        /transl_table=2
        /product="NADH dehydrogenase subunit 2"
        /protein_id="AAB87618.1"
CDS
  source
    1..1105
      /organism="Poecilia perugiae"
      /mitochondrion
      /db_xref="taxon:69232"
      /note="Tribe Poeciliini; Georgetown, Guyana"
      33..1079
        /codon_start=1
        /transl_table=2
        /product="NADH dehydrogenase subunit 2"
        /protein_id="AAB87621.1"
        /db_xref="PID:g2653597"
        /db_xref="GI:2653597"
        /translation="MAPPYISALIIISGLGTTMTFASTHWYLAWMGIEINTLAIPLM
        AONHPRATEATKYFFAQATASATLLFASISNAFLTGEWDITQPTPTSTLITLAL
        AMKIGLAPLHSMPEVMQGLNLTGLILSTWQKLAPLYLIYOQPNNTFTLGLLS
        IIVGGWGFNOVLKRLAYSSIAHLGWMILILSFSPPLALLTIYLLMTFSLFSF
        MLTRATHINSLTWAKIPILTYTPLLILSLGGLPLPGFMKWLILQELTKQDLAP
        LATLAALSLSLSLYFYRLSYAMTSLMPPNPNPAGTLPWRLNPQHNTFLMALTTSTIC
        LLPVTPAIMALLPF"
BASE COUNT 304 a 356 c 124 g 321 t
ORIGIN

Query Match 100.0%; Score 17; DB 4; Length 1105;
```

Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGACCAACTCCTAT 17
|||||
Db 230 CTCGACCAACTCCTAT 246

RESULT 4
AF084973 1105 bp DNA 27-APR-1999
LOCUS Heterandria formosa NADH dehydrogenase subunit 2 gene,
DEFINITION mitochondrial gene encoding mitochondrial protein, complete cds.
ACCESSION AF084973
NID 94689158
VERSION AF084973.1 GI:4689158
KEYWORDS Heterandria formosa.
SOURCE Mitochondrion Heterandria formosa
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Heterandria.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden.F., Ptacek.M.B., Rashed.M., Taphorn.D. and de
Figueiredo.C.A.
TITLE Molecular phylogeny of the live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae)
JOURNAL Mol. Phylogenet. Evol. (1999) In press
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden.F., Ptacek.M.B., Rashed.M. and de Figueiredo.C.A.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) Biological Sciences, Simon Fraser
University, Burnaby, BC V5A 1S6, Canada
FEATURES
Source
1. .1105
/organism="Heterandria formosa"
/mitochondrion
/db_xref="taxon:55104"
/country="USA:South Carolina, Savannah River"
<33. .>1079
/product="NADH dehydrogenase subunit 2"
33. .1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAD27788.1"
/db_xref="PID:94689159"
/db_xref="GI:4689159"
/translation="WAPIVHSTLITSLGTTLTFTASTHWYLAWMGIETILAIPLM
AGNHPRVEATKYFFAQATASATLLFAGASNAFLTGQWDILOTHPLPSTLVTLAL
ANKIGLPLHSMPEVMGVSLPTGLILSTWOKLAPLCILFOIQDPNPNIFITIGLS
IITGGWGFNQVQLRKILAYSIAHLGWMILILSFAPPLTIALFTYLTMTFSFSS
MLLRTTHSSSTWAKIPALTASVPLILSLGGLPPLTGLPKWLILQELTKQNLTL
IALSAALSLSFYFLRYLSYALMTMPNPPGLPWRLSSNNHTTIPALTTSATIL
LLPLAPAILAXFTF"

BASE COUNT 277 a 415 c 131 g 274 t 8 others
ORIGIN

Query Match 100.0%; Score 17; DB 4; Length 1105;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGACCAACTCCTAT 17
|||||
Db 230 CTCGACCAACTCCTAT 246

RESULT 5
185578 17 bp DNA 20-MAR-1998
LOCUS Heterandria formosa NADH dehydrogenase subunit 2 gene,
DEFINITION mitochondrial gene encoding mitochondrial protein, complete cds.
ACCESSION AF084973
NID 94689158
VERSION AF084973.1 GI:4689158
KEYWORDS Heterandria formosa.
SOURCE Mitochondrion Heterandria formosa
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Heterandria.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden.F., Ptacek.M.B., Rashed.M., Taphorn.D. and de
Figueiredo.C.A.
TITLE Molecular phylogeny of the live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae)
JOURNAL Mol. Phylogenet. Evol. (1999) In press
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden.F., Ptacek.M.B., Rashed.M. and de Figueiredo.C.A.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) Biological Sciences, Simon Fraser
University, Burnaby, BC V5A 1S6, Canada
FEATURES
Source
1. .1105
/organism="Heterandria formosa"
/mitochondrion
/db_xref="taxon:55104"
/country="USA:South Carolina, Savannah River"
<33. .>1079
/product="NADH dehydrogenase subunit 2"
33. .1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAD27788.1"
/db_xref="PID:94689159"
/db_xref="GI:4689159"
/translation="WAPIVHSTLITSLGTTLTFTASTHWYLAWMGIETILAIPLM
AGNHPRVEATKYFFAQATASATLLFAGASNAFLTGQWDILOTHPLPSTLVTLAL
ANKIGLPLHSMPEVMGVSLPTGLILSTWOKLAPLCILFOIQDPNPNIFITIGLS
IITGGWGFNQVQLRKILAYSIAHLGWMILILSFAPPLTIALFTYLTMTFSFSS
MLLRTTHSSSTWAKIPALTASVPLILSLGGLPPLTGLPKWLILQELTKQNLTL
IALSAALSLSFYFLRYLSYALMTMPNPPGLPWRLSSNNHTTIPALTTSATIL
LLPLAPAILAXFTF"

BASE COUNT 277 a 415 c 131 g 274 t 8 others
ORIGIN

Query Match 100.0%; Score 17; DB 4; Length 1105;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGACCAACTCCTAT 17
|||||
Db 230 CTCGACCAACTCCTAT 246

RESULT 5
185578 17 bp DNA 20-MAR-1998
LOCUS Heterandria formosa NADH dehydrogenase subunit 2 gene,
DEFINITION mitochondrial gene encoding mitochondrial protein, complete cds.
ACCESSION AF084973
NID 94689158
VERSION AF084973.1 GI:4689158
KEYWORDS Heterandria formosa.
SOURCE Mitochondrion Heterandria formosa
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Heterandria.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden.F., Ptacek.M.B., Rashed.M., Taphorn.D. and de
Figueiredo.C.A.
TITLE Molecular phylogeny of the live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae)
JOURNAL Mol. Phylogenet. Evol. (1999) In press
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden.F., Ptacek.M.B., Rashed.M. and de Figueiredo.C.A.
TITLE Direct Submission
JOURNAL Submitted (21-AUG-1998) Biological Sciences, Simon Fraser
University, Burnaby, BC V5A 1S6, Canada
FEATURES
Source
1. .1105
/organism="Heterandria formosa"
/mitochondrion
/db_xref="taxon:55104"
/country="USA:South Carolina, Savannah River"
<33. .>1079
/product="NADH dehydrogenase subunit 2"
33. .1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAD27788.1"
/db_xref="PID:94689159"
/db_xref="GI:4689159"
/translation="WAPIVHSTLITSLGTTLTFTASTHWYLAWMGIETILAIPLM
AGNHPRVEATKYFFAQATASATLLFAGASNAFLTGQWDILOTHPLPSTLVTLAL
ANKIGLPLHSMPEVMGVSLPTGLILSTWOKLAPLCILFOIQDPNPNIFITIGLS
IITGGWGFNQVQLRKILAYSIAHLGWMILILSFAPPLTIALFTYLTMTFSFSS
MLLRTTHSSSTWAKIPALTASVPLILSLGGLPPLTGLPKWLILQELTKQNLTL
IALSAALSLSFYFLRYLSYALMTMPNPPGLPWRLSSNNHTTIPALTTSATIL
LLPLAPAILAXFTF"

ACCESSION 185578
NID 93205296
VERSION 185578.1 GI:3205296
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 17)
AUTHORS Duff,G.W., Russell,G. and Eastell,R.
TITLE Detecting genetic predisposition for osteoporosis
JOURNAL Patent: US 5698399-A 1 16-DEC-1997;
FEATURES Location/Qualifiers
source
1. .17
/organism="unknown"
BASE COUNT 5 a 7 c 1 g 4 t
ORIGIN

Query Match 100.0%; Score 17; DB 5; Length 17;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCGACCAACTCCTAT 17
|||||
Db 1 CTCGACCAACTCCTAT 17

RESULT 6
HSILIRECA 12565 bp DNA 25-JUN-1997
LOCUS H sapiens gene for interleukin-1 receptor antagonist.
DEFINITION
ACCESSION X64532
NID 933798
VERSION X64532.1 GI:33798
KEYWORDS interleukin 1 alpha and beta homologue; interleukin 1 receptor
antagonist.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 12565)
AUTHORS Carrier,M.J.
TITLE Direct Submission
JOURNAL Submitted (17-FEB-1992) M.J. Carrier, Yamanouchi Research
Institute, Littlemore Hospital, Oxford, OX4 4XN, UK
REFERENCE 2 (bases 1 to 12565)
AUTHORS Lennard,A., Gorman,P., Carrier,M., Griffiths,S., Scotney,H.,
Sheer,D. and Solari,R.
TITLE Cloning and chromosome mapping of the human interleukin-1 receptor
antagonist gene
JOURNAL Cytokine 4 (2), 83-89 (1992)
MEDLINE 92338323
FEATURES Location/Qualifiers
source
1. .12565
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone_lib="lambda EMBL3 genomic library"
/map="q13-q14.1"
5949. .5953
5989. .6120
/number=1
Join(5989. .6120,7952. .8040,9418. .9530,11029. .12386)
Join(6005. .6120,7952. .8040,9418. .9530,11029. .11244)
/codon_start=1
/product="Interleukin-1 receptor antagonist"
/protein_id="CAA45832.1"
/db_xref="PID:g33799"
/db_xref="GI:33799"
/db_xref="SWISS-PROT:P18510"
/translation="MEICRGLRSLITLLFLFHSETICRPSGRKSKMQAFRIWDVN
QKTFYLRNNQLVAGYVLPVNVNLEKIDVPIEPHALFLGHHGKMLCSVKSGGDETR
LQLEAVNTDLSNRKQKRAFIKRSVGPITSPESACPGWFLCTAMEADQPVSLTN

```

MPDEGVMTKIFYFOEDE"
6121..7951
/number=1
7952..8040
/number=2
8041..9417
/number=2
9418..9530
/number=3
9531..11028
/number=3
11029..12386
/number=4
BASE COUNT 3217 a 2980 c 3072 g 3294 t 2 others
ORIGIN

Query Match 100.0%; Score 17; DB 10; Length 12565;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCAACACTCCTAT 17
Db 8868 CTCAGCAACACTCCTAT 8884

RESULT 7
LOCUS HSU65590 33414 bp DNA PRI 22-DEC-1997
DEFINITION Homo sapiens IL-1 receptor antagonist IL-1ra (IL-1RN) gene,
alternatively spliced forms, complete cds.
ACCESSION U65590
NID 92707374
VERSION U65590.1 GI:2707374
KEYWORDS
SOURCE human.
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 33414)
AUTHORS Jenkins,J.K., Drong,R.F., Shuck,M.E., Bienkowski,M.J.,
Slightom,J.L., Arend,W.P. and Smith,M.F. Jr.
TITLE Intracellular IL-1 receptor antagonist promoter: cell type-specific
and inducible regulatory regions
J. Immunol. 158 (2), 748-755 (1997)
JOURNAL 97146044
MEDLINE 2 (bases 1 to 33414)
AUTHORS Slightom,J.L.
REFERENCE Direct Submission
JOURNAL Submitted (30-JUL-1996) Molecular Biology, Pharmacia & Upjohn
Company, 301 Hennretti, Kalamazoo, MI 49007, USA
FEATURES
Source
1..33414
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="ic-1; P2q"
/chromosome="2"
/map="2q13-q14.1"
/dev_stage="adult"
repeat_region 1..236
/rpt_family="Alu"
/rpt_type-dispersed
repeat_region 5614..5910
/rpt_family="Alu"
/rpt_type-dispersed
repeat_region 7319..7590
/rpt_family="Alu"
/rpt_type-dispersed
mRNA
join(16075..16210,25876..25927,27756..27844,29053..29165,30664..32022)
/gene="IL-1RN"
/product="IL-1 receptor antagonist IL-1ra, short intracellular form"
gene
16075..32022
/genes="IL-1RN"
exon
16075..16210
/genes="IL-1RN"
note="exon icl; similar to GenBank Accession Number M55646"
join(16201..16210,25876..25927,27756..27844,29053..29165,30664..30879)
/genes="IL-1RN"
note="alternatively spliced form; similar to product encoded by GenBank Accession Number M55646"
/codon_start=1
/product="IL-1 receptor antagonist IL-1ra, short intracellular form"
protein_id="AA892268.1"
/db_xref="PID:g2707375"
/db_xref="GI:2707375"
translation="MALETICRPSGRKSSKMQAFRIWDVNQKTFYLRNNOLVAGYLOG
PNVNLKEDIQVPIEPHALFLGIHGKMLSCVKSQDETQLQLEAVNITLSENRRKOD
KREAFIRSDSGPTTSFSAACPGWFLCTAMEADQVSLTNMPDEGVMTKIFYFOEDE"
join(16201..16210,18246..18308,25876..25927,27756..27844,29053..29165,30664..30879)
/genes="IL-1RN"
note="alternatively spliced form"
/codon_start=1
/product="IL-1 receptor antagonist IL-1ra, long intracellular form"
protein_id="AA892269.1"
/db_xref="PID:g2707376"
/db_xref="GI:2707376"
translation="MALADLYEGGGGGEDNADSKETICRPSGRKSSKMQAFRIW
DVNQKTFYLRNNOLVAGYLOGPNVNLKEDIQVPIEPHALFLGIHGKMLSCVKSQD
ETQLQLEAVNITLSENRRKODKREAFIRSDSGPTTSFSAACPGWFLCTAMEADQVSL
TNMPDEGVMTKIFYFOEDE"
16211..25875
/genes="IL-1RN"
note="intron 1 of short intracellular form"
16211..18245
/genes="IL-1RN"
note="intron icl"
16398..17227
/rpt_family="Alu"
/rpt_type-dispersed
18246..18308
/genes="IL-1RN"
note="exon icl1; similar to GenBank Accession Number X84348"
18309..25779
/genes="IL-1RN"
note="intron icl1"
21027..21315
/rpt_family="Alu"
/rpt_type-dispersed
23996..24284
/rpt_family="Alu"
/rpt_type-dispersed
25780..25927
/genes="IL-1RN"
note="exon ls; similar to GenBank Accession Numbers X52015 and X53296"
join(25780..25927,27756..27844,29053..29165,30664..32022)
/genes="IL-1RN"
/product="IL-1 receptor antagonist IL-1ra, secreted form"
join(16075..16210,25876..25927,27756..27844,29053..29165,30664..30879)
/genes="IL-1RN"
note="alternatively spliced form"
/codon_start=1
/product="IL-1 receptor antagonist IL-1ra, secreted form"

```



```

/gene="fusa"
/codon_start=1
/transl_table=11
/product="elongation factor EF-G"
/protein_id="AAC06402.1"
/db_xref="PID:g2982776"
/db_xref="GI:2982776"
/translation="MAREVPIEKLRNIGIVAHIDAGKTTTTERILYITGTYKIGEVH
EGATMDMPOEKERGITITVATTACYWTRNGERYOINIIDTPGHVDFSEVVVRSMKV
LDGIVLFSVAGEVQOSEANWRAQFQVPRIAFINKMDRLGADRYRFEKELEKLT
IKPVAIQIPGADQDEQEGVIDLMMKRAIRWLEETIGAKYEVVDIPPEYQEKAGWEK
MIETIVDELMKLEGEIISIDELRKARLKERLVPVLCGSAFKNGQVPLL
DAVIDLPSIDIPPVKGTNPKTGEVHRPSDDEPCAYAFKVMSPDPYAGQYIIRV
FSTLKGASVYNATKDEKORAGRLILMHANSREEIQOVSAGEICAVVGLDAATGTL
CDKHPIILEKLEFPDPVISMATEPKTKDQELSOVLNKFEMKEDPTERTATDPETGO
ILIHGMLHLEIMVDMKREYGLIEVNVKPOVAYKETIRKKAIGEGKFIKQGVGRGO
YGHAIIELEPPGAGFEFIDDIHGVPKREFIPSVKEKGKAMQNGILAGYPPVDVR
VRLFDGYSDESSDAFQVAGSLAFKAAKADPVLLEPIMEVEVETPEKYGVGVIG
DLNSRRKGIMGNKGVTIVIKAHVPLAEMFGYATTLRSLTQGRGTFINKFSHYDEVP
QQTAEKIIGERMAGKSS"
2117. 3334
/gene="tufA1"
/notes="aq_005"
2117. 3334
/gene="tufA1"
/codon_start=1
/transl_table=11
/product="elongation factor EF-Tu"
/protein_id="AAC06403.1"
/db_xref="PID:g2982777"
/db_xref="GI:2982777"
/translation="MAKEKPERTKEHVNVGTIGHVDHGKSTLTSAITCVLAAGLVEGG
KACFKYEIDKAPEKERGITINITHVEYETAKRYAHVDCPCGHADYIKNMITGAQ
MDGAILVSAADQPMQTRRHVLLAQOVNPYIVFMNKCDDWDDLELLELEVEVRE
LLSKYEPGDEVIVRGSALGALOELEQNSPGKWEISIKELLNADDEIYIPTPOREVDK
PFJLMEIDPEVSGRGTVTGRYERGLRPGDEVEIVGLRLEPKMTVTSIEMFRKVL
DEALPDGNIIVLLRGVKDDVERGOVLAQPGSVKAHRFRAQVYVLSKEEGGRHTPF
VNYRPOFYFTADVGTGVVVKLPEGVEMVPGDNVELEVELEIAPVALEEGLRFAIRBGG
RTVAGGVTKILD"
3346. 3660
/gene="rpsJ"
/notes="aq_008"
3346. 3660
/gene="rpsJ"
/codon_start=1
/transl_table=11
/product="ribosomal protein S10"
/protein_id="AAC06399.1"
/db_xref="PID:g2982773"
/db_xref="GI:2982773"
/translation="MEQEKIRIKLAYDHRLLDOSVKQIIEYVKRTGGVYVKGPIPLPT
RKKWCVLSPHFDQSRHFEIRFSRIIDIRFIPQTIELMELSLPAGVDVEVKM
RG"
3665. 4390
/gene="rplC"
/notes="aq_009"
3665. 4390
/gene="rplC"
/codon_start=1
/transl_table=11
/product="ribosomal protein L03"
/protein_id="AAC06393.1"
/db_xref="PID:g2982767"
/db_xref="GI:2982767"
/translation="MELGLIGKGVMTRVLLKDGTALPVTVIKFPVYVYVQVKSQNTK
DGNALQIGAYEAKERHLTKPLIGHKKKGVPLLRLWFRWNPPEFQSGQOLKYVED
VFAPGLVDVWGSKRGFGAKMRWDFAFPKSHGHYHRAVGAQCQRTDPRGVWKG
KRMPGHYGAKPRVQGLFVYVDVLPFEENAILVKGSPGHNGKIVVVEQSTIAYRKSQL
KOKRYOFIITENLAKEESQEAES"
4387. 4986
/gene="rplD"
/notes="aq_011"
4387. 4986
/gene="fusa"
/codon_start=1
/transl_table=11
/product="elongation factor EF-G"
/protein_id="AAC06402.1"
/db_xref="PID:g2982776"
/db_xref="GI:2982776"
/translation="MAREVPIEKLRNIGIVAHIDAGKTTTTERILYITGTYKIGEVH
EGATMDMPOEKERGITITVATTACYWTRNGERYOINIIDTPGHVDFSEVVVRSMKV
LDGIVLFSVAGEVQOSEANWRAQFQVPRIAFINKMDRLGADRYRFEKELEKLT
IKPVAIQIPGADQDEQEGVIDLMMKRAIRWLEETIGAKYEVVDIPPEYQEKAGWEK
MIETIVDELMKLEGEIISIDELRKARLKERLVPVLCGSAFKNGQVPLL
DAVIDLPSIDIPPVKGTNPKTGEVHRPSDDEPCAYAFKVMSPDPYAGQYIIRV
FSTLKGASVYNATKDEKORAGRLILMHANSREEIQOVSAGEICAVVGLDAATGTL
CDKHPIILEKLEFPDPVISMATEPKTKDQELSOVLNKFEMKEDPTERTATDPETGO
ILIHGMLHLEIMVDMKREYGLIEVNVKPOVAYKETIRKKAIGEGKFIKQGVGRGO
YGHAIIELEPPGAGFEFIDDIHGVPKREFIPSVKEKGKAMQNGILAGYPPVDVR
VRLFDGYSDESSDAFQVAGSLAFKAAKADPVLLEPIMEVEVETPEKYGVGVIG
DLNSRRKGIMGNKGVTIVIKAHVPLAEMFGYATTLRSLTQGRGTFINKFSHYDEVP
QQTAEKIIGERMAGKSS"
2117. 3334
/gene="tufA1"
/notes="aq_005"
2117. 3334
/gene="tufA1"
/codon_start=1
/transl_table=11
/product="ribosomal protein L04"
/protein_id="AAC06394.1"
/db_xref="PID:g2982768"
/db_xref="GI:2982768"
/translation="MKIGDVEVEDDVNVKKVHVWVVKK..AKPROGTHSTKTRG
EVAYSRRKILPQKGTGNARHGERGVNIFVGGGVANHPKPRDYETPLPKRVKVLGLKMA
LSDKARNNAIIFVDNIDLGKKPKTKAIEFLKNLGVEKEKVLVVIPEKADVLKXSPFN
LPNVRVLLPEGLNVYDVLWANKLVIQDCLDRIYKVEA"
4990. 5301
/gene="rplW"
/notes="aq_012"
4990. 5301
/gene="rplW"
/codon_start=1
/transl_table=11
/product="ribosomal protein L23"
/protein_id="AAC06397.1"
/db_xref="PID:g2982771"
/db_xref="GI:2982771"
/translation="MSQRKPWEIIRPIITEKSNRLMEDYNKYTFEVALDASKPEIKY
AVEKLEFNVKVKVNTMIVKPKKKRVNKKFQYGTTRKKWKAIVTLEKGHKIDILEPQA
K"
5313. 6227
/gene="rplB"
/notes="aq_013"
5313. 6227
/gene="rplB"
/codon_start=1
/transl_table=11
/product="ribosomal protein L02"
/protein_id="AAC06392.1"
/db_xref="PID:g2982766"
/db_xref="GI:2982766"
/translation="MGVRKLKPVNTNCTRHAVLYDFEIEKLVKKGKEWLLKKNQVEP
EKSLLLWHRAKGRSQRQNRNITARRGGGHHKLYRIIDFARDSLVPFAKVVSLEYDPF
RSARICLLHYADGERKRYIIWPEGLKVGDTVMSWEDAEAKPLPEIKPGNANPLKYI
PEGTIVHNIEFIPGKGQIARAAGTWAQVLRSTRKGYLVLRMPSPSEVRMIHERCNAT
VGRVGLAEHELKVGKAGRARWLGRPRTGTAMNPVDHPHGGEGEGRTRCKHPESPWG
WTKGYKTRRGKKYSQDFIVTRDRGRPL"
6340. 6900
/gene="rpsS"
/notes="aq_015"
6340. 6900
/gene="rpsS"
/codon_start=1
/transl_table=11
/product="ribosomal protein S19"
/protein_id="AAC06401.1"
/db_xref="PID:g2982775"
/db_xref="GI:2982775"
/translation="MREAIKRYGSFELLKGNLTPLEALLEERKLVLENLKKOLKEA
HKGKPKIEAGDEKELIREVYNKAQAEVRALEIIVNVRKRYELIYAQYKQWTEKKAY
VDPKLWVRIRKKNNEGERKVVRTYSRATTIIEFVGHITIAVHGKTFVPVYITQDMVG
HKLGEFAPRTFKGHPKTAKEVVKKK"
7018. 7314
/gene="rplV"
/notes="aq_016a"
7018. 7314
/gene="rplV"
/codon_start=1
/transl_table=11
/product="ribosomal protein L22"
/protein_id="AAC06396.1"
/db_xref="PID:g2982770"
/db_xref="GI:2982770"
/translation="MFSGKFRKWTPTPTCFPAVIPKKAARIAEKVLKSVIANAEQKG
LDLDRLYIKKAVADDGPILKKWIPRAHGRATMVRKRLSHITIVLEEKPEGKEE"
7317. 7955
/gene="rpsC"
/notes="aq_017"
7317. 7955
```



```
gene      MFLVLANGWELVVLVLSVRSYQ"
complement(5640..6635)
/gene="aq_1922"
CDS      complement(5640..6635)
/gene="aq_1922"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC07721.1"
/db_xref="PID:g2984189"
/db_xref="GI:2984189"
/translation="MKFEWVWVEPDRKIVSVALEGANNAVVIPEKGRVVEEVKKVGRI
TVAPDGLKLGEDVYVVLKGEDEERAAYPPNVKIVETIDWTVIPLENLIAQRE
ELYAVYKNAQAEVALOTLEKGVKGVLSKRDINEIKKVGQIVSSEHEKLELVTKIT
KILPLGLCDRCVCDIISLLHRCGMLVGNSSGGMELVHAETENPVVAARPRVNAGA
VHMIIRPNRTKYICELKAGDKVMVYDKGRGRVTVYGRKAVRVERPMLLIIEGRNK
KLSCLIQNAETIRLTKPDGTPIVSSELKEGDEVLGVEEAGRHFGMKVEETIEEK"
6660..7550
/gene="aq_1923"
CDS      6660..7550
/gene="aq_1923"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC07719.1"
/db_xref="PID:g2984187"
/db_xref="GI:2984187"
/translation="MKKALGILAILLILGVYFAYDKYMDNKAKEQVEYFLDKLRKS
GKSKYVYKFIGEILIKYVYDRNGREEFKIEIIEKLESETEGKFLFNKAPLK
VKGGLLEYKDFKFNLFVSEAKPREKFEHLRSLSDYDPEAEVNFISFILGNVDH
APKWTALSDRPEEVSFVLSGELSKINSLEVYVROKGFERVIRKKEAQRKTPE
EFKELIRIEEKLKAESEFERNLLDAPEKLEKGEIKVYVKNPPLKIQDLFVVA
AVOKDERELLKLLNPVIEVK"
complement(7524..8213)
/gene="aq_1924"
CDS      complement(7524..8213)
/gene="aq_1924"
/codon_start=1
/transl_table=11
/product="putative protein"
/protein_id="AAC07720.1"
/db_xref="PID:g2984188"
/db_xref="GI:2984188"
/translation="MFLVVDVGNSSVDIALWEGKVKVDFLKLSSHEEFLEKFEFPLKA
LGISVQSSEKVRGKPIKFKKENFPIDYKTPETLGTDRVALAYSARKFYGKN
VVVISAGTALVDLVEGKFGFTLGLGKKLKLSDLAEGIPFEVEFIFLGRS
TRCVLGAGYARESTEFIKSTLKRKVKRKFKVITGGEGYFSKFGIYDPLVHRG
MRNLLYLXRI"
complement(8210..9775)
/gene="hyuB"
CDS      complement(8210..9775)
/note="aq_1925"
/gene="hyuB"
/codon_start=1
/transl_table=11
/product="N-methylhydantoinase B"
/protein_id="AAC07716.1"
/db_xref="PID:g2984184"
/db_xref="GI:2984184"
/translation="MNSSILLEVFNKAFSSIAEEMGLVLQRTAFSPNIKERDFSCAV
FDKGNLVAQAHI PVHLGSMPLSVKEAIKAFGEFEGDMVVLNDPVMGTHLPDIT
LVAPFYKELLFFLANRAHADVGATPGSMPLSTSIQEGFIIPPLKLMRGELNE
DFMKVLRNVRPEEREGDFKQIMANITGIKRLRELIEREGIEKLVFFSKELMNYSE
RIIRERKALPEGEYEFEDMEDDGYRKDVKIHKLKVEKELIDFDSKDDQTRGG
INAVRAITLSAVYVAILSGSDIPVNEGCFRPLEVITRKGVVDAEFPAAGVAGNVE
TSQIVDVVLGAFSAIPKVPAAASOGTMNHVTIGGINPENGEPFTYYETIGGMCAS
AKGDSVASHSHMTNLTNPIDALEHYVPILVTEYSIRKSGSGEGMHRGSDGIVREY
FLDVSVTLSERRRLAPGLFGGKPGMPGINLIKTEEGEKMPKSFYSYLKGGDLR
VETPGGGGYMKEE"
9830..9902
/product="trRNA-Thr"
trRNA    /anticodon="(pos:9863)..9865,aa:Thr)
9925..10009
```

```
/product="trRNA-Tyr"
/anticodon="(pos:9959)..9961,aa:Tyr)
10019..10091
/product="trRNA-Gly"
/anticodon="(pos:10052)..10054,aa:Gly)
10124..10196
/product="trRNA-Thr"
/anticodon="(pos:10157)..10159,aa:Thr)
10282..11499
/gene="tufA2"
/note="aq_1928"
10282..11499

Query Match      90.6%; Score 15.4; DB 2; Length 11891;
Best Local Similarity 94.1%; Pred. No. 87;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 CTCAGCAACTCTCTAT 17
        ||||| |||||
Db      11150 CTCAGCAGCACTCTAT 11134

RESULT 11
ECU86689
LOCUS   ECU86689 3360 bp DNA BCT 17-MAY-1999
DEFINITION Erwinia chrysanthemi exuT (exuT) and exuR (exuR) genes, complete cds.
ACCESSION U86689 U35884
NID       94826585
VERSION   U86689.1 GI:4826585
KEYWORDS . Erwinia chrysanthemi.
SOURCE    Erwinia chrysanthemi.
ORGANISM  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Erwinia.
REFERENCE 1 (bases 1 to 1820)
AUTHORS   Haseloff,B.J., Freeman,T.L., Valmeekam,V., Melkus,M.W., Oner,F., Valachovic,M.S. and San Francisco,M.J.
TITLE     The exuT gene of Erwinia chrysanthemi EC16: nucleotide sequence, expression, localization, and relevance of the gene product
JOURNAL   Mol. Plant Microbe Interact. 11 (4), 270-276 (1998)
MEDLINE   98192001
REFERENCE 2 (bases 1821 to 3360)
AUTHORS   Valmeekam,V., Freeman,T.L., Loh,Y.-L. and San Francisco,M.J.D.
TITLE     Regulation of galacturonate transport in Erwinia chrysanthemi EC16
JOURNAL   Unpublished
REFERENCE 3 (bases 1 to 3360)
AUTHORS   San Francisco,M.J.D.
TITLE     Direct Submission
JOURNAL   Submitted (11-SEP-1995) Biological Sciences, Texas Tech University, Flint and Main, Lubbock, TX 79409-3131, USA
COMMENT   On May 14, 1999 this sequence version replaced gi:2429101.
FEATURES
         Location/Qualifiers
             1..3360
                /organism="Erwinia chrysanthemi"
                /strain="EC16"
                /db_xref="taxon:556"
                296..1333
                   /gene="exuT"
                296..1333
                   /gene="exuT"
                /codon_start=1
                /transl_table=11
                /evidence=experimental
                /product="exuT"
                /protein_id="AAB70881.1"
                /db_xref="PID:g1794192"
                /db_xref="GI:1794192"
                /translation="MFKIKGLRWYMGIVTIGTLVLRNAIAAAAPTLOQLHIST
                QQYSYIIAAYSACYTIMQPVAGYVLDVLTGKVGAMFALWALFCAGTALANSWGGLA
                VARGAVGMAEAMIPAGKASSEVPKASERSVAVGVFNVGSSIGGMALPPLVVAIMA
                HSWOMAFILITGALSILVWALCWLIFYKHPKDOKLSTEREVILSQEQHQHGNAKRM
                SAWQILNRQFNQIALPFLAEPANGTINAWIPMFYKAYGFENKIAMFAMPLFA
```

gene DLGCIILGYMPLFQYKYNLIVSRKLVVTLGALLMIGTGITGLFTSPYVAIACCA
AALPTSPCPVR"
1957. 2736
/gene="exur"
CDS 1957. 2736
/gene="exur"
/function="galacturonic acid uptake"
/codon_start=1
/transl_table=11
/evidence-experimental
/product="exur"
/protein_id="AAD30185.1"
/db_xref="PID:g4826586"
/db_xref="GI:4826586"
/translation="MALTESRRLYQOLAAELKORINGVYPVGDKLPAERNISEEMNV
SRIVREAIIMLEVEGYEVKRGSGIHVISNQORHLLAANGSDADFLTAGPFELLOAR
OLIESIAFAATQVTRDIIQIMEIOENARQEDRFROSEWDLKHFVQVALATONSAM
ATIVERKWSQRHNPYKRLHEHIDKSIESCEDHDIILKALIRKDPYAAKLAMWQH
LENTKMLFRATTDDEFNVDRYMEAFENPVVHLDHTNGKL"

BASE COUNT 811 a 917 c 879 g 753 t
ORIGIN

Query Match 90.6%; Score 15.4; DB 2; Length 3360;
Best Local Similarity 94.1%; Pred. No. 1e+02; Indels 0; Gaps 0;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGCAACTCCTAT 17
||||||| |||||
Db 426 CTCAGCAACTCCTAT 442

RESULT 12
AF031387 1105 bp DNA VRT 30-NOV-1997
LOCUS Poecilia vivipara NADH dehydrogenase subunit 2 gene, mitochondrial
DEFINITION gene encoding mitochondrial protein, complete cds.
ACCESSION AF031387
NID 92653578
VERSION AF031387.1 GI:2653578
KEYWORDS Poecilia vivipara.
SOURCE Mitochondrion Poecilia vivipara
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Poecilia.

REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden,F., Ptacek,M., Rashed,M., Taphorn,D. and de Figueiredo,C.A.
TITLE Molecular phylogeny of a live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae) based on NADH dehydrogenase
subunit 2 sequence variation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden,F., Ptacek,M. and Rahsed,M.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1997) Biological Sciences, Simon Fraser
University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
LOCATION/Qualifiers 1. .1105

FEATURES
source /organism="Poecilia vivipara"
/mitochondrion
/db_xref="taxon:69236"
/note="Tribe Poeciliini"
33. .1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAB87612.1"
/db_xref="PID:g2653579"
/db_xref="GI:2653579"
/translation="MAPPIYSALIIISLGLTMTFASTHWYLAWMGIEINTLAIPLM
AQNHIPRAIEATTKYFFVQATASATLLFASISNAFLTGWDIINTPHTLTSTLTIAL

CDS
1. .1105
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAB87612.1"
/db_xref="PID:g2653579"
/db_xref="GI:2653579"
/translation="MAPPIYSALIIISLGLTMTFASTHWYLAWMGIEINTLAIPLM
AQNHIPRAIEATTKYFFVQATASATLLFASISNAFLTGWDIINTPHTLTSTLTIAL

AMKIGLAPLHSMPEVMOGLSLITGLTSTWOKLAPLYLIYOIQPNPNIFITLGLLS
IIIGWGFNQVQLRKILAYSSIAHGLWMILILSFSPALLITLIYILMTFSLSFSF
MLTRTHINSLSITWAKIPILTVSTPLILSLGGXPPLTGFPKPKHILQELKQNLAP
LATLAALSLSLSYFLRLSYAMTILMPNPNFAGMLPWRNLNPHRNTFLTATTSTIC
LLPMPAIMALEPFF"

BASE COUNT 318 a 375 c 115 g 296 t 1 others
ORIGIN

Query Match 90.6%; Score 15.4; DB 4; Length 1105;
Best Local Similarity 94.1%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGCAACTCCTAT 17
||||||| |||||
Db 230 CTCAGCAACTCCTAT 246

RESULT 13
AF031388 1105 bp DNA VRT 30-NOV-1997
LOCUS Poecilia gillii NADH dehydrogenase subunit 2 gene, mitochondrial
DEFINITION gene encoding mitochondrial protein, complete cds.
ACCESSION AF031388
NID 92653580
VERSION AF031388.1 GI:2653580
KEYWORDS Poecilia gillii.
SOURCE Mitochondrion Poecilia gillii
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Poecilia.

REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden,F., Ptacek,M., Rashed,M., Taphorn,D. and de Figueiredo,C.A.
TITLE Molecular phylogeny of a live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae) based on NADH dehydrogenase
subunit 2 sequence variation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden,F., Ptacek,M. and Rahsed,M.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1997) Biological Sciences, Simon Fraser
University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
LOCATION/Qualifiers 1. .1105

FEATURES
source /organism="Poecilia gillii"
/mitochondrion
/db_xref="taxon:69226"
/note="Tribe Poeciliini"
33. .1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAB87613.1"
/db_xref="PID:g2653581"
/db_xref="GI:2653581"
/translation="MAPPIYSALIIISLGLTMTFASTHWYLAWMGIEINTLAIPLM
AMKIGLAPLHSMPEVMOGLSLITGLTSTWOKLAPLYLIYOIQPNPNIFITLGLLS
IIIGWGFNQVQLRKILAYSSIAHGLWMILILSFSPALLITLIYILMTFSLSFSF
MLTRTHINSLSITWAKIPILTVSTPLILSLGGXPPLTGFPKPKHILQELKQNLAP
LATLAALSLSLSYFLRLSYAMTILMPNPNRAGTLFWRXNPRHKSUPLAXTTTSTIC
LLPLTPAIMSLMPX"

BASE COUNT 318 a 354 c 119 g 310 t 4 others
ORIGIN

Query Match 90.6%; Score 15.4; DB 4; Length 1105;
Best Local Similarity 94.1%; Pred. No. 1.2e+02; Indels 0; Gaps 0;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGCAACTCCTAT 17

```
|||||
Db 230 CTCAGCAACTTCTAT 246

RESULT 14
AF031390
LOCUS
DEFINITION Poecilia sphenops NADH dehydrogenase subunit 2 gene, mitochondrial
ACCESSION AF031390
NID q2653584
VERSION AF031390.1 GI:2653584
KEYWORDS Poecilia sphenops.
SOURCE Mitochondrion Poecilia sphenops
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Poecilia.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden, F., Ptacek, M., Rashed, M., Taphorn, D. and de Figueiredo, C.A.
TITLE Molecular phylogeny of a live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae) based on NADH dehydrogenase
subunit 2 sequence variation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden, F., Ptacek, M. and Rahsed, M.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1997) Biological Sciences, Simon Fraser
University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
FEATURES
source
1..1105
/organism="Poecilia reticulata"
/mitochondrion
/db_xref="taxon:8081"
/notes="Tribe Poeciliini; Aripo River, Caroni Drainage,
Trinidad"
33..1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAB87615.1"
/db_xref="PID:q2653593"
/db_xref="GI:2653593"
/translation="MAPPYISALIIISLGLTMTFASTHWYLAWMGIEINTLAIPLM
SONHTPRATEATTKYFFAQATASATLLFASISNAFLTGWMDILOTPNTLTSSLTITLAL
AMKIGLAPLHSHWPEVMQGLNLTGLISTWOKLAPLYLIYQIOPNPNDFIALGILS
ILVGGWGFNOVLKRLIAYSSNAHLGWMILILSPPLALTIYIILMTSLFSF
MLTRTHINSLATWAKVPIITISTPLVLLSGGLPPLTGFMFKWILLOELTKQDLAP
LATMAALSLSLFYLRLSYAMTLTMTPPNPAGTLPWRLNPRHNTLPMALTTTSTIC
LLPVTPTMSLLPF"
BASE COUNT 319 a 367 c 123 g 295 t 1 others
ORIGIN

Query Match 90.6%; Score 15.4; DB 4; Length 1105;
Best Local Similarity 94.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGCAACTCCTAT 17
|||||
Db 230 CTCAGCAACTTCTAT 246

RESULT 15
AF031394
LOCUS
DEFINITION Poecilia reticulata NADH dehydrogenase subunit 2 gene,
mitochondrial gene encoding mitochondrial protein, complete cds.
ACCESSION AF031394
NID q2653592
VERSION AF031394.1 GI:2653592
KEYWORDS Poecilia reticulata.
SOURCE guppy.
Mitochondrion Poecilia reticulata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Poecilia.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden, F., Ptacek, M., Rashed, M., Taphorn, D. and de Figueiredo, C.A.
TITLE Molecular phylogeny of a live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae) based on NADH dehydrogenase
subunit 2 sequence variation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden, F., Ptacek, M. and Rahsed, M.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1997) Biological Sciences, Simon Fraser
University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
FEATURES
source
1..1105
/organism="Poecilia sphenops"
/mitochondrion
/db_xref="taxon:69235"
/notes="Tribe Poeciliini"
33..1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAB87615.1"
/db_xref="PID:q2653585"
/db_xref="GI:2653585"
/translation="MAPPYISALIIISLGLTMTFASTHWYLAWMGIEINTLAIPLM
AQNHTPRATEATTKYFFAQATASATLLFASISNAFLTGWMDILOTPNTLTSSLTITLAL
AMKIGLAPLHSHWPEVMQGLNLTGLISTWOKLAPLYLIYQIOPNPNDFIALGILS
ILVGGWGFNOVLKRLIAYSSNAHLGWMILILSPPLALTIYIILMTSLFSF
MLTRTHINSLATWAKVPIITISTPLVLLSGGLPPLTGFMFKWILLOELTKQDLAP
LATMAALSLSLFYLRLSYAMTLTMTPPNPAGTLPWRLNPRHNTLPMALTTTSTIC
LLPVTPTMSLLPF"
BASE COUNT 319 a 367 c 123 g 295 t 1 others
ORIGIN

Query Match 90.6%; Score 15.4; DB 4; Length 1105;
Best Local Similarity 94.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGCAACTCCTAT 17
|||||
Db 230 CTCAGCAACTTCTAT 246

RESULT 15
AF031394
LOCUS
DEFINITION Poecilia reticulata NADH dehydrogenase subunit 2 gene,
mitochondrial gene encoding mitochondrial protein, complete cds.
ACCESSION AF031394
NID q2653592
VERSION AF031394.1 GI:2653592
KEYWORDS Poecilia reticulata.
SOURCE guppy.
Mitochondrion Poecilia reticulata
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;
Neopterygii; Teleostei; Euteleostei; Acanthopterygii;
Atherinomorpha; Cyprinodontiformes; Cyprinodontidae; Poeciliidae;
Poecilia.
REFERENCE 1 (bases 1 to 1105)
AUTHORS Breden, F., Ptacek, M., Rashed, M., Taphorn, D. and de Figueiredo, C.A.
TITLE Molecular phylogeny of a live-bearing fish genus Poecilia
(Cyprinodontiformes: Poeciliidae) based on NADH dehydrogenase
subunit 2 sequence variation
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1105)
AUTHORS Breden, F., Ptacek, M. and Rahsed, M.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1997) Biological Sciences, Simon Fraser
University, 8888 University Drive, Burnaby, BC V5A 1S6, Canada
FEATURES
source
1..1105
/organism="Poecilia reticulata"
/mitochondrion
/db_xref="taxon:8081"
/notes="Tribe Poeciliini; Aripo River, Caroni Drainage,
Trinidad"
33..1079
/codon_start=1
/transl_table=2
/product="NADH dehydrogenase subunit 2"
/protein_id="AAB87615.1"
/db_xref="PID:q2653593"
/db_xref="GI:2653593"
/translation="MAPPYISALIIISLGLTMTFASTHWYLAWMGIEINTLAIPLM
SONHTPRATEATTKYFFAQATASATLLFASISNAFLTGWMDILOTPNTLTSSLTITLAL
AMKIGLAPLHSHWPEVMQGLNLTGLISTWOKLAPLYLIYQIOPNPNDFIALGILS
ILVGGWGFNOVLKRLIAYSSNAHLGWMILILSPPLALTIYIILMTSLFSF
MLTRTHINSLATWAKVPIITISTPLVLLSGGLPPLTGFMFKWILLOELTKQDLAP
LATMAALSLSLFYLRLSYAMTLTMTPPNPAGTLPWRLNPRHNTLPMALTTTSTIC
LLPVTPTMSLLPF"
BASE COUNT 332 a 341 c 128 g 304 t
ORIGIN

Query Match 90.6%; Score 15.4; DB 4; Length 1105;
Best Local Similarity 94.1%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CTCAGCAACTCCTAT 17
|||||
Db 230 CTCAGCAACTACTAT 246

Search completed: September 18, 1999, 15:49:01
Job time: 1421 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 05:27:20 ; Search time 213.04 Seconds
(without alignments)
19.965 Million cell updates/sec

Title: US-09-037-472-5
Perfect score: 17
Sequence: 1 CTCAGCAACACTCCTAT 17

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	17	100.0	17	T89947	Human Interleukin-
2	17	100.0	17	T89409	Human IL-1RN gene
3	17	100.0	17	V32393	Interleukin-1RN pr
4	17	100.0	17	V60230	PCR primer used to
5	17	100.0	17	V62390	IL-1 receptor anta
6	17	100.0	12565	V62392	Human interleukin-
7	17	100.0	17	X16621	Interleukin 1 (441
8	14.4	84.7	369	T72131	CEA-specific antibody
9	14.4	84.7	786	X05452	CEA6 antibody sing
10	14	82.4	2481	V40019	Mouse Pax6 gene. U
11	14	82.4	2481	V40035	Mouse Pax6 gene SE
12	14	82.4	2245	V61485	Human secreted pro
13	14	82.4	3597	V81414	Human prenylcystei
14	14	82.4	2084	X37379	Human secreted pro
15	14	82.4	2078	X37439	Human secreted pro
16	13.8	81.2	1220	Q05864	Candida tropicalis
17	13.8	81.2	906	Q05865	Candida tropicalis
18	13.8	81.2	305	Q59719	Human brain expres
19	13.8	81.2	2111	V24142	Homo sapiens BARD1
20	13.8	81.2	2531	V24115	Homo sapiens BARD1
21	13.8	81.2	2531	V24116	Homo sapiens BARD1
22	13.8	81.2	2531	V24117	Homo sapiens BARD1
23	13.8	81.2	2531	V24118	Homo sapiens BARD1
24	13.8	81.2	2510	V24119	Homo sapiens BARD1
25	13.8	81.2	2531	V24120	Homo sapiens BARD1
26	13.8	81.2	2531	V24121	Homo sapiens BARD1
27	13.8	81.2	2531	V24122	Homo sapiens BARD1
28	13.8	81.2	2531	V24123	Homo sapiens BARD1
29	13.8	81.2	2531	V24124	Homo sapiens BARD1
30	13.8	81.2	2531	V24125	Homo sapiens BARD1
31	13.8	81.2	610	V43261	DNA encoding a hum
32	13.8	81.2	931	V59584	Human secreted pro
33	13.8	81.2	4344	V65381	Arabidopsis phyto
34	13.8	81.2	2254	V65382	Arabidopsis phyto
35	13.8	81.2	3200	X33947	Human HCMV inducib
36	13.4	78.8	7399	Q11816	RNA encoding enter
37	13.4	78.8	2382	Q10238	Encodes membrane e
38	13.4	78.8	2082	Q05789	Sequence comprisin
39	13.4	78.8	3547	T93746	TURF 2H3 DNA fragm
40	13.4	78.8	1568	T94534	Alfalfa cinnanoyl
41	13.4	78.8	1829	V15914	Coxsackievirus CVB
42	13.4	78.8	5630	V22647	Drosophila melanoga
43	13.4	78.8	7313	V29268	Nucleotide sequenc

c 44 13.4 78.8 1497 1 V48110 Nucleotide sequenc
45 13.4 78.8 1558 1 V84453 Human secreted pro

ALIGNMENTS

RESULT 1

T89947
ID T89947 standard; DNA; 17 BP.
AC T89947, 1998 (first entry)
DT 05-MAR-1998
DE Human Interleukin-1 receptor antagonist intron 2 PCR primer 1.
KW Interleukin-1 receptor antagonist; IL-1ra; ulcerative colitis;
KW diagnosis; prognosis; inflammatory bowel disease; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9725445-A1.
PD 17-JUL-1997.
PF 08-JAN-1997; U00042.
PR 12-JAN-1996; US-587911.
PA (CEDA-) CEDARS SINAI MEDICAL CENT.
PA (UYVI-) UNIV VIRGINIA PATENT FOUND.
PI Cominelli F, Pizarro T, Rottler JI, Yang H;
WPI; 97-372880/34.
DR Screening for ulcerative colitis in subjects of Jewish ancestry - by
PT detecting allele 2 of the VNTR (variable number of tandem repeats)
PT polymorphism at intron 2 of the IL-1 receptor antagonist gene
Claim 7; Page 17; 22pp; English.
PS This PCR primer and primer T89948 are used to amplify intron 2 of the
CC human interleukin-1 receptor antagonist gene (IL-1ra) in a novel method
CC to screen for ulcerative colitis (UC) in a subject of Jewish ancestry.
CC There is an association between allele 2 of the variable number of tandem
CC repeats (VNTR) polymorphism at intron 2 of IL-1ra, an important
CC endogenous regulator of inflammation, and UC in humans of Jewish
CC ancestry. This method can be used for the diagnosis and prognosis of UC
CC in Jewish patients for UC and distinguishing UC from Crohn's disease (CD)
CC and other inflammatory disease of the bowel.
SQ Sequence 17 BP; 5 A; 7 C; 1 G; 4 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCAGCAACACTCCTAT 17

|||||

Db 1 CTCAGCAACACTCCTAT 17

RESULT 2

T89409

ID T89409 standard; cDNA; 17 BP.

AC T89409;

DT 22-APR-1998 (first entry)

DE Human IL-1RN gene intron 2 PCR primer 1.

KW Osteoporosis; Interleukin-1 receptor antagonist; IL-1RN; allele;

KW bone mineral density; post-menopause; PCR primer; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9738135-A1.

PD 16-OCT-1997.

PF 03-APR-1997; U05626.

PR 05-APR-1996; US-628282.

PA (MEDI-) MEDICAL SCI SYSTEMS INC.

PI Duff GW, Eastell R, Russell G;

WPI; 97-512741/47.

DR Detecting genetic predisposition for osteoporosis - by detecting

PT Interleukin-1 receptor antagonist gene IL-1RN allele 2 in the

PT genomic DNA of a patient

PS Claim 2; Page 9; 21pp; English.

CC PCR primers T89409 and T89410 are used to amplify a region of the

CC interleukin-1 receptor antagonist gene, IL-1RN, intron 2 which contains

CC a variable number tandem repeat (VNTR) region that gives rise to five
 CC alleles. This product is used for predicting the risk of osteoporosis in
 CC a subject by determining the allelic and genetic polymorphism pattern for
 CC IL-1RN in genomic DNA. A pattern of at least one copy of the IL-1RN
 CC allele 2 indicates an increased susceptibility to osteoporosis. The
 CC methods can predict low bone mineral density (BMD) and the rate of bone
 CC density loss and thereby a susceptibility to osteoporosis. Individuals
 CC so identified can then be treated more aggressively to prevent or retard
 CC the occurrence of disease.
 SQ Sequence 17 BP; 5 A; 7 C; 1 G; 4 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCACACTCCTAT 17
 Db 1 CTCAGCACACTCCTAT 17
 |||||

RESULT 3

V32393
 ID V32393 standard: DNA; 17 BP.
 AC V32393;
 DT 11-SEP-1998 (first entry)
 DE Interleukin-1RN primer 1.
 KW Genetic polymorphism; PCR; primer: amplification: interleukin-1RN;
 KW sight threatening diabetic retinopathy; interleukin-1-alpha; IL-1RN;
 KW interleukin-1-beta; ss.
 OS Synthetic.
 OS Homo sapiens.
 PN WO9815653-A1.
 PD 16-APR-1998.
 PF 09-OCT-1997; G02790.
 PR 10-OCT-1996; GB-021129.
 PA (DUFF) DUFF G.
 PA (RENN) RENNIE I.
 PA (RICH) RICHARDSON R.
 PI Duff G, Rennie I, Richardson R;
 WPI: 98-240835/21.

PT Predicting increased risk of sight-threatening diabetic retinopathy
 PI - comprises identifying genetic polymorphism pattern for genes
 PI IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
 PT symptoms occur

PS Claim 2; Page 33; 41pp; English.
 CC Interleukin-1RN (IL-1RN) primers 1 and 2 (V32394) were used to amplify
 CC the IL-1RN gene region to identify polymorphism of the VNTR region
 CC at the IL-1RN intron 2 locus. The invention claims to provide a method
 CC for predicting the risk of sight threatening diabetic retinopathy. The
 CC method involves isolating DNA from a patient and determining the DNA
 CC polymorphism pattern of the genes that code for interleukin-1-alpha,
 CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
 CC identified is then compared with controls of known DNA polymorphism
 CC patterns thereby identifying patients carrying a genetic polymorphism
 CC associated with increased risk of sight threatening diabetic retinopathy.
 CC The method may be able to identify diabetic patients at risk before the
 CC clinically detectable disorders occur. Polymorphism pattern
 CC determination of IL genes involved PCR reactions using primers
 CC V32389-V32398. The method is also claimed to be useful in conjunction
 CC with identification of other genes associated with sight threatening
 CC diabetic retinopathy in genomic DNA and therefore, in identifying
 CC diabetic patients expressing multiple risk patterns.
 SQ Sequence 17 BP; 5 A; 7 C; 1 G; 4 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCACACTCCTAT 17
 Db 1 CTCAGCACACTCCTAT 17
 |||||

RESULT 4
 V60230
 ID V60230 standard: DNA; 17 BP.
 AC V60230;
 DT 25-NOV-1998 (first entry)
 DE PCR primer used to amplify interleukin-1 receptor antagonist (IL-1RN).
 KW Interleukin-1 receptor antagonist; IL-1RN; predisposition;
 KW coronary artery disease; screen; PCR primer; ss.
 OS Synthetic.
 PN WO9840517-A1.
 PD 17-SEP-1998.
 PF 09-MAR-1998; U04725
 PR 10-MAR-1997; US-813456.
 PA (MED-1) MEDICAL SCI SYSTEMS INC.
 PI Crossman DC, Duff GW, Francis SE;
 DR WPI: 98-520829/44.
 PT Detection of predisposition to coronary artery disease - by
 PT comparative measurement of levels of expression of alleles from the
 PT interleukin 1 locus
 PS Claim 6; Page 15; 22pp; English.
 CC PCR primers V60230-31 were used to amplify alleles associated with
 CC the interleukin-1 receptor antagonist (IL-1RN). The specification
 CC describes a method for determination of a patient's predisposition
 CC to coronary artery disease. The method comprises comparing an
 CC allele with a second allele which is predictive of coronary artery
 CC disease, where similarity between the first and second alleles
 CC indicates a predisposition to coronary artery disease. The method is
 CC used to genotype an individual's interleukin (IL)-1 loci, the
 CC overexpression of which correlates with coronary artery disease. The
 CC method is used to screen a patient's predisposition to coronary
 CC artery disease.
 SQ Sequence 17 BP; 5 A; 7 C; 1 G; 4 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
 Best Local Similarity 100.0%; Pred. No. 1.3;
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCACACTCCTAT 17
 Db 1 CTCAGCACACTCCTAT 17
 |||||

RESULT 5
 V62390
 ID V62390 standard: DNA; 17 BP.
 AC V62390;
 DT 19-JAN-1999 (first entry)
 DE IL-1 receptor antagonist gene intron 2 PCR primer #1.
 KW Human; interleukin-1 receptor antagonist gene; IL-1; polymorphism;
 KW diagnosis; osteoporosis; PCR primer; ss.
 OS Synthetic.
 OS Homo sapiens.
 PN WO9844150-A1.
 PD 08-OCT-1998.
 PF 27-MAR-1998; G00944.
 PR 27-MAR-1997; GB-006359.
 PA (GEMI-) GEMINI RES LTD.
 PI Keen RW, Specter TD;
 DR WPI: 98-557135/47.
 PT Diagnosis of osteoporosis by determining genotype of interleukin-1
 PT receptor antagonist gene - useful for diagnosing patient
 PT pre-disposition or susceptibility to osteoporosis and for
 PT therapeutic intervention
 PS Claim 9; Page 10; 36pp; English.
 CC A method has been developed for the diagnosis of osteoporosis comprising
 CC determining the genotype of an interleukin-1 (IL-1) receptor antagonist
 CC gene (IL-1RN). The present sequence represents a PCR primer adapted to
 CC amplify a portion of intron 2 of an IL-1RN for use in the method of the
 CC invention. The method can be used for the diagnosis of disease,
 CC including diagnosis of osteoporosis and predisposition or susceptibility

```

CC to osteoporosis and for therapy.
SQ Sequence 17 BP; 5 A; 7 C; 1 G; 4 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCAACACTCCTAT 17
   |||||
DB 1 CTCAGCAACACTCCTAT 17

RESULT 6
ID V62392 standard; DNA; 12565 BP.
AC V62392;
DE Human interleukin-1 receptor antagonist gene.
KW Human; interleukin-1 receptor antagonist gene; IL-1; polymorphism;
OS diagnosis; osteoporosis; ds.
PN Homo sapiens.
PD WO9844150-A1.
PD 08-OCT-1998.
PF 27-MAR-1998; G00944.
PR 27-MAR-1997; GB-006359.
PA (GEMI-) GEMINI RES LTD.
PI Keen RW, Spector TD;
DR WPI; 98-557195/47.
PT Diagnosis of osteoporosis by determining genotype of interleukin-1
PT receptor antagonist gene - useful for diagnosing patient
PT pre-disposition or susceptibility to osteoporosis and for
PT therapeutic intervention
PS Disclosure; Page 21-27; 36pp; English.
CC A method has been developed for the diagnosis of osteoporosis comprising
CC determining the genotype of an interleukin-1 (IL-1) receptor antagonist
CC gene (IL-1RN). The present sequence represents the human interleukin-1
CC receptor antagonist gene. The method can be used for the diagnosis of
CC disease, including diagnosis of osteoporosis and predisposition or
CC susceptibility to osteoporosis and for therapy.
SQ Sequence 12565 BP; 3217 A; 2980 C; 3072 G; 3294 T;

Query Match 100.0%; Score 17; DB 1; Length 12565;
Best Local Similarity 100.0%; Pred. No. 2.7;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCAACACTCCTAT 17
   |||||
DB 8868 CTCAGCAACACTCCTAT 8864

RESULT 7
ID X16621
AC X16621;
DE 29-APR-1999 (first entry)
DE Interleukin 1 (44112332) haplotype PCR primer #15.
KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;
KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;
KW Graves disease; systemic lupus erythematosus; lichen sclerosis;
KW ulcerative colitis; PCR primer; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9854359-A1.
PD 03-DEC-1998.
PD 21-MAY-1998; G01481.
PR 29-MAY-1997; GB-011040.
PA (CAMP/) CAMP N J.
PA (COXA/) COX A.
PA (DGIO/) DE GIOVINE F S.
PA (DUFF/) DUFF G.
PI Camp NJ, Cox A, De Giovine FS, Duff G;

```

```

DR WPI; 99-080814/07.
PT New method of determining a patient's susceptibility to inflammatory
PT disorders - by detecting the presence of an IL-1 (44112332)
PT haplotype, useful in designing treatment strategies that modulate
PT the activity of proteins produced by the IL-1 gene cluster
PS Claim 3; Page 33; 49pp; English.
CC A method has been developed for determining a patient's susceptibility
CC to an inflammatory disorder. The method comprises the detection of an
CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the
CC patient, where its presence indicates susceptibility to an inflammatory
CC disorder. X16607 to X16631 represent PCR primer used in the method for
CC detecting the IL-1 (44112332) haplotype. The method provides kits for
CC the early prediction of a patient's susceptibility to inflammatory
CC disorders, including coronary artery disease, osteoporosis, nephropathy
CC in diabetes mellitus, alopecia areata, Graves disease, systemic lupus
CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of
CC alleles of the haplotype can be applied to particular inflammatory
CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,
CC psoriasis, and insulin dependent diabetes. The identification of a
CC disease-associated haplotype enables determination of which alleles are
CC causative, and this information is useful in designing treatment
CC strategies, including gene therapy and treatment using various agents
CC that modulate the activity of proteins produced by the IL-1 gene cluster.
CC Some alleles from the IL-1 gene cluster are associated with particular
CC inflammatory diseases, and insufficient IL-1 production appears to act
CC centrally in the pathology of these diseases. Therefore, the use of IL-1
CC gene clusters is useful in determining genetic susceptibility to
CC inflammatory diseases, including those with a multifactorial etiology
CC with a polygenic component.
SQ Sequence 17 BP; 5 A; 7 C; 1 G; 4 T;

Query Match 100.0%; Score 17; DB 1; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCAACACTCCTAT 17
   |||||
DB 1 CTCAGCAACACTCCTAT 17

RESULT 8
ID T72131 standard; DNA; 369 BP.
AC T72131;
DE 07-DEC-1997 (first entry)
DE CEA-specific antibody CEA6 VH gene.
DE Carcinoembryonic antigen; CEA; human; antibody; scFv;
KW tumour marker; lung cancer; breast cancer; colon cancer;
KW adenocarcinoma; diagnosis; ss.
PN Homo sapiens.
PN WO9720932-A1.
PD 12-JUN-1997.
PD 09-DEC-1996; G03043.
PF 11-OCT-1996; GB-021295.
PR 07-DEC-1995; GB-025004.
PR 23-MAY-1996; GB-010824.
PA (CAMP-) CAMBRIDGE ANTIBODY TECHNOLOGY.
PA Allen DJ, McCafferty JG, Osbourn JK;
DR WPI; 97-319779/29.
DR P-PSDB; W19881.
PT Specific binding members for human carcinoembryonic antigen - bind
PT to the A3-B3 extracellular domain of hCEA and are substantially
PT non-cross-reactive with human liver cells; used for diagnosing
PT cancer.
PS Example 1; Fig 1a; 128pp; English.
CC This nucleotide sequence codes for the heavy chain variable region
CC (VH) (W19881) of human carcinoembryonic antigen (hCEA)-specific
CC antibody CEA6. VH (T72126-32) and VL (T72133-35) gene sequences
CC were determined for anti-hCEA antibodies CEA1-CEA7 (see W19876-85)
CC that had been obtained by selection from a universal phage display
CC library. A claimed specific binding member (A) comprises an hCEA
CC specific antibody antigen binding domain that has a dissociation

```



```
CC constant for hCEA of less than 1 x 10-8 M, is non-cross-reactive
CC with human liver cells, and preferentially binds to the A3-B3
CC extracellular domain of hCEA and/or to cell-associated hCEA over
CC hCEA over soluble hCEA. Preferred (A) include pairings of VH and
CC VL sequences from CEA1-7, or their CDR sequences, as well as CEA6
CC VH and VL variants. (A) is used to detect cells expressing hCEA,
CC in vivo or in vitro, especially tumour cells for diagnosing cancer,
CC e.g. adenocarcinoma of the colon, lung or breast.
SQ Sequence 369 BP: 85 A; 101 C; 109 G; 74 T;

Query Match 84.7%; Score 14.4; DB 1; Length 369;
Best Local Similarity 93.8%; Pred. No. 45;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCAGCAACTCTCTAT 17
DB 86 TCAGCAACTCTCTAT 101
||||||| |||||

RESULT 9
X05452
AC X05452: standard; DNA: 786 BP.
DE CEA6 antibody single-chain Fv (scFv) fragment encoding DNA.
KW Trimeric polypeptide; tetranectin trimerising structural element; TTSE;
KW fusion protein; ligand binding structure; toxin; enzyme; cytokine; CEA6;
KW artificial antibody; pharmacokinetic; pharmacodynamic; gene therapy;
KW transfection; imaging; tumour; human; tetranectin; ss.
OS Unidentified.
PN W09856906-A1.
PD 17-DEC-1998.
PF 11-JUN-1998; DK0245.
PR 11-JUN-1997; DK-000685.
PA (ETVE/) ETZERODT M.
PA (GRAV/) GRAVERSEN N J H.
PA (HOLT/) HOLTET T L.
PA (KAST/) KASTROP J S.
PI Etzerodt M, Graversen NJH, Holtet TL, Kastrup JS,
PI Larsen IK, Nielsen BB, Thøgersen HC;
DR WPI: 99-080897/07.
PT New monomer polypeptide constructs for diagnosis and therapy -
PT comprise a tetranectin trimerising structural element covalently
PT linked to at least one heterologous moiety for providing functional
PT activity
PS Example 4; Page 63-64; 110pp; English.
CC The invention relates to the design of trimeric polypeptides using
CC polypeptide structural elements derived from the tetranectin protein
CC family. The trimeric polypeptides constructed as a monomer polypeptide
CC construct comprise at least one tetranectin trimerising structural
CC element (TTSE) which is covalently linked to at least one heterologous
CC moiety, the TTSE being capable of forming a stable complex with 2 other
CC TTSEs, with the proviso that the heterologous moiety is different from
CC any of the fusion proteins C1H6FXTN123, H6FXTN12, H6FXTN23
CC (W94261 to W94264). The TTSE can be used for the construction of
CC conjugates with heterologous moieties such as a ligand binding structure,
CC a toxin, a detectable label, an in situ activatable substance, an enzyme,
CC a radioactive moiety, a cytokine, a non-proteinaceous polymer, a photo
CC cross-linking agent, or a group facilitating conjugation of the monomer
CC polypeptide construct to a target. They can be used as vehicles for
CC assembling antibody fragments into oligomeric or multivalent entities
CC for generating chimeric artificial antibodies having preselected
CC pharmacokinetic and/or pharmacodynamic properties. The constructs can
CC be used for targeted gene therapy involving selective delivery of the
CC material for transfection or infection of the specific population of
CC cells. They can also be used for delivering a substance to a cell or
CC tissue or for delivering an imaging or toxin-conjugated antibody to a
CC tumour. They can also be used for prevention or treating a disease or
CC for diagnosis. The TTSE provides a stable structure which can act as a
CC vehicle for a wide variety of conjugates. The present sequence represents
CC a nucleotide sequence encoding a CEA6 antibody single-chain (scFv)
CC fragment. This is used in the construction of trimerised and hexamerised
```

```
CC scFv antibodies.
SQ Sequence 786 BP: 187 A; 211 C; 227 G; 161 T;

Query Match 84.7%; Score 14.4; DB 1; Length 786;
Best Local Similarity 93.8%; Pred. No. 49;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCAGCAACTCTCTAT 17
DB 86 TCAGCAACTCTCTAT 101
||||||| |||||

RESULT 10
V40019
ID V40019 standard; cDNA; 2481 BP.
AC V40019:
DT 12-OCT-1998 (first entry)
DE Mouse Pax6 gene.
KW Mouse; Pax6; Pax4; pancreatic cell; differentiation status;
KW developmental status; transgenic mammal; hypoglycaemia; diabetes;
KW neuronal disorder; tumour; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 163..1473
FT /tag= a
FT /product= "Pax6"
PN W09829565-A2.
PD 09-JUL-1998.
PF 30-DEC-1997; E07320.
PR 31-DEC-1996; US-778394.
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
PI Gruss P, St-Onge L;
DR WPI: 98-388143/33.
DR P-PSDB: W69509.
PT Use of Pax6 nucleic acids and proteins - useful for, e.g. developing
PT products for diagnosis, prevention and treatment of hypoglycaemia,
PT diabetes, neuronal disorders and tumours
PS Claim 23; Page 38-42; 66pp; English.
CC A method has been developed for testing the developmental status in
CC pancreatic cells (PCs) of a mammal comprising: (a) determining the level
CC or status of Pax6 mRNA in PCs of the mammal; and/or (b) determining the
CC level or status of Pax6 protein in PCs of the mammal; and (c) comparing
CC the level or status of Pax6 mRNA and/or Pax6 protein with the
CC corresponding level in normal PCs. The present sequence encodes Pax6
CC protein. The method can be used for determining the developmental status
CC of PCs as indicative of hypoglycaemia, diabetes, neuronal disorders or
CC tumours. The products can be used for developing agents for treating or
CC preventing such disorders.
CC Sequence 2481 BP: 743 A; 554 C; 571 G; 613 T;

Query Match 82.4%; Score 14; DB 1; Length 2481;
Best Local Similarity 100.0%; Pred. No. 92;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGCAACTCTCTA 16
DB 1038 CAGCAACTCTCTA 1051
||||||| |||||

RESULT 11
V40035
ID V40035 standard; cDNA; 2481 BP.
AC V40035:
DT 12-OCT-1998 (first entry)
DE Mouse Pax6 gene SEQ ID NO:3.
KW Mouse; Pax4; Pax6; pancreatic cell; differentiation status; tumour;
KW developmental status; transgenic mammal; diabetes; neuronal disorder; ss.
OS Mus sp.
FH Key Location/Qualifiers
FT CDS 163..1473
FT /tag= a
```

FT W09829566-A2. /product= "Pax6"
 PN 09-JUL-1998.
 PD 30-DEC-1997; E07321.
 PF 31-DEC-1996; US-778423.
 PR (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.
 PA Gruss P. Sosa-Pineda B;
 PI WPI: 98-388144/33.
 DR P-PSDB; W69513.
 DT Use of Pax4 nucleic acids and proteins - useful for, e.g. developing
 PT products for diagnosis, prevention and treatment of diabetes,
 PT neuronal disorders and tumours
 PS Claim 23: Page 43-47; 70pp; English.
 CC A method has been developed for testing the developmental status in
 CC pancreatic cells (PC's) of a mammal comprising: (a) determining the
 CC level or status of Pax4 mRNA in PC's of the mammal; and/or (b)
 CC determining the level or status of Pax4 protein in PC's of the mammal,
 CC and (c) comparing the level or status of Pax4 mRNA and/or Pax4 protein
 CC with the corresponding level in normal PC's. The present invention also
 CC describes a nucleic acid sequence encoding a functional and expressible
 CC Pax4 protein and optionally a second nucleic acid sequence encoding a
 CC functional and expressible Pax6 protein, for the preparation of a
 CC therapeutic composition for treating, preventing and/or delaying
 CC diabetes and/or a neuronal disorder in a mammal. The present sequence
 CC encodes a Pax6 protein. The method can be used for determining the
 CC development of PC's as indicative of diabetes, neuronal disorders or
 CC tumours. The products can be used for developing agents for treating
 CC these disorders.
 SQ Sequence 2481 BP; 743 A; 554 C; 571 G; 613 T;

Query Match 82.4%; Score 14; DB 1; Length 2481;
 Best Local Similarity 100.0%; Pred. No. 92;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGCAACACTCTCTA 16
 DB 1038 CAGCAACACTCTCTA 1051
 RESULT 12
 V61485/C
 ID V61485; standard; cDNA; 2245 BP.
 AC 11-JAN-1999 (first entry)
 DT Human secreted protein do568_l1 cDNA.
 DE Secreted protein; human; do568_l1; ds.
 KW Homo sapiens.
 OS
 FH Key Location/Qualifiers
 FT CDS 359..1369
 FT /*tag= a
 PN W09841539-A2.
 PD 24-SEP-1998.
 PF 19-MAR-1998; U05474.
 PR 18-MAR-1998; US-040963.
 PA 19-MAR-1997; US-820493.
 FA (GEMV) GENETICS INST INC.
 PI Agostino MJ, Jacobs K, LaVallie ER, McCoy JM, Merberg D,
 PI Racine LA, Spaulding V, Treacy M;
 DR WPI: 98-521163/44.
 DR P-PSDB; W79094.
 DT New polynucleotide(s) encoding secreted human proteins - derived
 PT from human foetal kidney, adult testes and adult or foetal brain
 PT cDNA libraries
 PT Claim 30; Page 82-84; 112pp; English.
 CC This full-length cDNA clone, designated do568_l1, codes for a novel
 CC secreted human protein (see W79094). It was isolated from a human
 CC adult testis cDNA library using methods which are selective for
 CC cDNAs encoding secreted proteins, or was identified as encoding
 CC a secreted or transmembrane protein on the basis of computer
 CC analysis of the encoding protein. The nucleotide sequence shows
 CC homology to some database sequences. The invention provides
 CC cDNA clones (see V61477-87) from human foetal kidney, adult testis,

CC and adult or foetal brain cDNA libraries that code for secreted
 CC proteins (see W79087-97). These clones are deposited as
 CC ATCC 98364. The polynucleotides and proteins are predicted to have
 CC useful biological activities which would make them suitable for
 CC treating, preventing or ameliorating medical conditions in humans
 CC and animals, although no supporting data is given. Suggested
 CC activities include nutritional, immune stimulating (e.g. as
 CC vaccines) or suppressing, haematopoiesis regulating, tissue growth,
 CC activin/inhibin, chemotactic/chemokinetic, haemostatic and
 CC thrombolytic, receptor/ligand, antiinflammatory, cadherin/tumour
 CC invasion suppressor and tumour inhibition activities. The
 CC polynucleotides are also stated to be useful for gene therapy.
 SQ Sequence 2245 BP; 443 A; 632 C; 617 G; 546 T;

Query Match 82.4%; Score 14; DB 1; Length 2245;
 Best Local Similarity 100.0%; Pred. No. 91;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 CAGCAACACTCTCTA 16
 DB 1429 CAGCAACACTCTCTA 1416
 RESULT 13
 V81414
 ID V81414 standard; cDNA; 3597 BP.
 AC 12-APR-1999 (first entry)
 DT Human prenylcysteine carboxyl methyltransferase cDNA.
 DE Human prenylcysteine carboxyl methyltransferase; PCMT; human;
 KW prenylcysteine carboxyl methyltransferase; PCMT; human;
 KW inhibitor; screening; inflammation; cancer; cell proliferation;
 KW psoriasis; Ras protein; therapy; ss.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT CDS 44..898
 FT /*tag= a
 PN W09856924-A1.
 PD 17-DEC-1998.
 PF 11-JUN-1998; U12328.
 PR 11-JUN-1997; US-049304.
 PA (UVNY) UNIV NEW YORK STATE.
 PI Phillips M;
 DR WPI: 99-070326/06.
 DR P-PSDB; W67730.
 DT New mammalian prenylcysteine carboxyl methyltransferase - and its
 PT inhibitors, useful for treating inflammation, cancer and other
 PT proliferative diseases
 PS Claim 12; Page 67-70; 95pp; English.
 CC This cDNA clone codes for a new human prenylcysteine carboxyl
 CC methyltransferase (PCMT, see W67730) involved in post-translational
 CC modification of prenylated proteins. It was isolated following
 CC an EST database search for mammalian sequences related to the
 CC Saccharomyces cerevisiae isoprenylcysteine methyltransferase Stel4
 CC gene. PCR primers (see V81415-16) based on an isolated EST were
 CC used in a PCR amplification of human HL60 cell cDNA. A PCR product
 CC (see V81417) was used to screen an HL60 cDNA library to isolate
 CC the human PCMT cDNA. Homology of the expression product to
 CC Stel4 was 26%. Also new are an expression vector containing DNA
 CC encoding human PCMT, hosts transformed with this vector, compounds
 CC that include an antigen-binding part of an antibody (Ab) specific
 CC for PCMT; and a method for screening for, and isolating,
 CC inhibitors of PCMT. The inhibitors are used to treat inflammation,
 CC cancer and other hyperproliferative diseases such as psoriasis, e.g.
 CC by inhibiting prenylation of Ras proteins. PCMT is also used as a
 CC reagent for carboxy methylation of substrates, e.g. to produce
 CC functional Ras proteins, and for raising Ab. When over-expressed
 CC in a cell, peptide fragments of PCMT may act as inhibitors.
 CC Fragments of PCMT nucleic acids are used as probes to isolate
 CC related sequences from other sources, and nucleic acids
 CC expressing a dominant negative variant of PCMT can be used in gene
 CC therapy.
 SQ Sequence 3597 BP; 876 A; 873 C; 880 G; 968 T;

Mon Sep 20 14:16:57 1999

us-09-037-472-5.rng

page 7

Db 1236 CAGCAACTCCTA 1223

Search completed: September 18, 1999, 05:27:21
Job time: 1686 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:37 : Search time 1405 Seconds
(without alignments)
23.867 Million cell updates/sec

Title: US-09-037-472-5
Perfect score: 17
Sequence: 1 CTCAGCAACTCTCTAT 17

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17	100.0	393	21	T71181	T71181 yc52d04.r1
2	15	88.2	540	28	AA102335	AA102335 z191b05.s
3	15	88.2	433	28	AA122310	AA122310 z188a06.s
4	15	88.2	483	29	AA135360	AA135360 z025e03.s
5	15	88.2	495	35	AA587215	AA587215 nm82a06.s
6	15	88.2	385	35	C28914	C28914 C28914 Rice
7	15	88.2	438	44	AI290208	AI290208 q179g10.x
8	14.4	84.7	366	20	D34140	D34140 CELK041E9R
9	14.4	84.7	235	25	N58629	N58629 vv57e12.r1
10	14.4	84.7	563	26	W75479	W75479 me56a02.r1
11	14.4	84.7	629	38	AA391390	AA391390 LD10274.S
12	14.4	84.7	2374	39	AF038351	AF038351 AF038251
13	14.4	84.7	438	41	AI040382	AI040382 ox16c10.x
14	14.4	84.7	551	44	AI257823	AI257823 LP06306.S
15	14.4	84.7	633	44	AI258185	AI258185 LP01318.S
16	14.4	84.7	570	45	AI353647	AI353647 zeh0805.s
17	14.4	84.7	443	47	AI477924	AI477924 fb49b05.x
18	14.4	84.7	503	51	AI723320	AI723320 fc35g01.x
19	14	82.4	415	20	D23826	D23826 RICR0273A.R
20	14	82.4	363	20	T30256	T30256 EST13578.Hu
21	14	82.4	331	20	Z24956	Z24956 HSB75H092.S
22	14	82.4	290	20	Z44501	Z44501 HSC22B051.n
23	14	82.4	416	21	T77786	T77786 yd20b11.r1
24	14	82.4	485	23	H47299	H47299 yp77f10.r1
25	14	82.4	422	23	H47705	H47705 yp75h05.r1
26	14	82.4	707	26	W27616	W27616 35c4 Human
27	14	82.4	386	27	AA043334	AA043334 zk53c09.r
28	14	82.4	473	28	AA102312	AA102312 z191b05.r
29	14	82.4	415	28	AA121086	AA121086 zm22d03.r
30	14	82.4	385	30	AA232934	AA232934 zr46d03.s
31	14	82.4	504	33	AA415740	AA415740 vd27g07.s
32	14	82.4	507	40	AA933081	AA933081 om85f07.s
33	14	82.4	464	43	AI198187	AI198187 q151f05.x
34	14	82.4	603	46	AI402823	AI402823 GH22123.S
35	14	82.4	339	47	AI467191	AI467191 vd27g08.x
36	14	82.4	256	47	AI490486	AI490486 EST249020
37	14	82.4	426	53	HSM006824	HSM006824
38	13.8	81.2	347	20	D33538	AI041974 Homo sapi
39	13.8	81.2	291	20	D34472	D33538 CELK032D7R
40	13.8	81.2	403	20	D34568	D34472 CELK046D4R
41	13.8	81.2	360	20	D35963	D34568 CELK008G8R
42	13.8	81.2	305	20	M78523	D35963 CELK027E5F
43	13.8	81.2	349	20	T24016	M78523 EST00871.Fe
44	13.8	81.2	573	21	T67556	T24016 seq2172.3HF
45	13.8	81.2	702	54	HSM011713	T67556 yd12h07.s1
						AI046863 Homo sapi

ALIGNMENTS

RESULT 1

T71181
LOCUS T71181 393 bp mRNA
DEFINITION yc52d04.r1 Stratagene liver (#937224) Homo sapiens cDNA clone
IMAGE:84295 5', mRNA sequence.
ACCESSION T71181
NID 9685702
VERSION T71181.1 GI:685702

01-MAR-1995

```

KEYWORDS      EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 303)
               Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
               Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
               Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
               Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
               Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
               Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
               and Marra, M.
TITLE          Generation and analysis of 280,000 human expressed sequence tags
JOURNAL        Genome Res. 6 (9), 807-828 (1996)
MEDLINE        97044478
COMMENT        On Sep 21, 1992 this sequence version replaced gi:276333.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 736
High quality sequence stops: 306 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL; contact the
Consortium (info@image.llnl.gov) for further information.
Insert Length: 736 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 306.
Location/Qualifiers
1..393
/organism="Homo sapiens"
/db_xref="GDB:501352"
/db_xref="taxon:9606"
/clone="IMAGE:84295"
/clone_lib="Stratagene liver (#937224)"
/sex="male"
/dev_stage="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: liver; Vector: pBluescript SK; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. Hepatectomy from normal male caucasian. Average
insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3'"

BASE COUNT    107 a 80 c 114 g 86 t 6 others
ORIGIN
|||||
1 CTCAGCAACTCTCTAT 17
Db 49 CTCAGCAACTCTCTAT 65

Query Match      100.0%; Score 17; DB 21; Length 393;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCAGCAACTCTCTAT 17
Db 49 CTCAGCAACTCTCTAT 65

RESULT 2
LOCUS      AA102335 540 bp mRNA EST 31-JUL-1997
DEFINITION 219b05.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
IMAGE:511953 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION  AA102335
NID         91647042
VERSION     AA102335.1 GI:1647042
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

KEYWORDS      EST.
SOURCE         human.
ORGANISM       Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
AUTHORS        Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 540)
               Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
               Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
               Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
               Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
               Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
               Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
               and Marra, M.
TITLE          Generation and analysis of 280,000 human expressed sequence tags
JOURNAL        Genome Res. 6 (9), 807-828 (1996)
MEDLINE        97044478
COMMENT        On Sep 21, 1992 this sequence version replaced gi:276333.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 736
High quality sequence stops: 306 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL; contact the
Consortium (info@image.llnl.gov) for further information.
Insert Length: 736 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 306.
Location/Qualifiers
1..540
/organism="Homo sapiens"
/db_xref="GDB:384448"
/db_xref="taxon:9606"
/clone="IMAGE:511953"
/clone_lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3'"

BASE COUNT    128 a 139 c 137 g 132 t 4 others
ORIGIN
|||||
1 CTCAGCAACTCTCT 15
Db 402 CTCAGCAACTCTCT 388

Query Match      88.2%; Score 15; DB 28; Length 540;
Best Local Similarity 100.0%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCAGCAACTCTCT 15
Db 402 CTCAGCAACTCTCT 388

RESULT 3
LOCUS      AA122310/c 433 bp mRNA EST 19-NOV-1996
DEFINITION 218a06.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
IMAGE:511666 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION  AA122310
NID         91678669
VERSION     AA122310.1 GI:1678669
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
               1 (bases 1 to 433)
               Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
               Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
               Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
               Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
               Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
               Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
               and Marra, M.
TITLE          Generation and analysis of 280,000 human expressed sequence tags
JOURNAL        Genome Res. 6 (9), 807-828 (1996)
MEDLINE        97044478
COMMENT        On Sep 21, 1992 this sequence version replaced gi:276333.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 736
High quality sequence stops: 306 Source: IMAGE Consortium, LLNL This
clone is available royalty-free through LLNL; contact the
Consortium (info@image.llnl.gov) for further information.
Insert Length: 736 Std Error: 0.00
Seq primer: M13RP1
High quality sequence stop: 306.
Location/Qualifiers
1..393
/organism="Homo sapiens"
/db_xref="GDB:501352"
/db_xref="taxon:9606"
/clone="IMAGE:84295"
/clone_lib="Stratagene liver (#937224)"
/sex="male"
/dev_stage="49 years old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: liver; Vector: pBluescript SK; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dr. Hepatectomy from normal male caucasian. Average
insert size: 1.1 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3'"

BASE COUNT    107 a 80 c 114 g 86 t 6 others
ORIGIN
|||||
1 CTCAGCAACTCTCTAT 17
Db 49 CTCAGCAACTCTCTAT 65

RESULT 2
LOCUS      AA102335/c 540 bp mRNA EST 31-JUL-1997
DEFINITION 219b05.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
IMAGE:511953 3' similar to contains Alu repetitive element;; mRNA
sequence.
ACCESSION  AA102335
NID         91647042
VERSION     AA102335.1 GI:1647042
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

```

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On May 18, 1995 this sequence version replaced gi:811366.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -40M13 fwd. from Amersham
High quality sequence stop: 371.

FEATURES

source

Location/Qualifiers
1..433
/organism="Homo sapiens"
/db_xref="GDB:3844161"
/db_xref="taxon:9606"
/clone="IMAGE:511666"
/clone_lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3'"

BASE COUNT 98 a 112 c 108 g 111 t 4 others

ORIGIN

Query Match 88.2%; Score 15; DB 28; Length 433;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCAACACTCT 15

Db 403 CTCAGCAACACTCT 389

RESULT 4

AAL35360/c

LOCUS

DEFINITION 483 bp mRNA EST 06-AUG-1997
2028e03.s1 Stratagene colon (#937204) Homo sapiens cDNA clone
IMAGE:587932 3' similar to contains Alu repetitive element;; mRNA
sequence.

ACCESSION

AAL35360

NID

91696464

VERSION

AAL35360.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 483)

Hillier, L., Lennon, G., Becker, M., Donald, M.F., Chiapelli, B.,

Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,

Rohlfing, T., Schellenger, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,

and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1326680.

Contact: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@wustl.wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

Insert Length: 2139 Std Error: 0.00

Seq primer: -40M13 fwd. from Amersham

High quality sequence stop: 81.

FEATURES

source

Location/Qualifiers
1..483
/organism="Homo sapiens"
/db_xref="GDB:4620234"
/db_xref="taxon:9606"
/clone="IMAGE:587932"
/clone_lib="Stratagene colon (#937204)"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: colon; Vector: pBluescript SK-; Site_1:
ECORI; Site_2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. T-84 colonic epithelial cell line. Average
insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3'"

BASE COUNT 116 a 122 c 119 g 122 t 4 others

ORIGIN

Query Match 88.2%; Score 15; DB 29; Length 433;

Best Local Similarity 100.0%; Pred. No. 1.6e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTCAGCAACACTCT 15

Db 398 CTCAGCAACACTCT 384

RESULT 5

AA587215/c

LOCUS

DEFINITION

495 bp mRNA EST

26-SEP-1997

nn82a06.s1 NCI-CGAP-Co9 Homo sapiens cDNA clone IMAGE:1090354 3'

similar to contains Alu repetitive element;; mRNA sequence.

ACCESSION

AA587215

NID

92398029

VERSION

AA587215.1

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 495)

AUTHORS

NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

Unpublished (1997)

JOURNAL

On Sep 1, 1995 this sequence version replaced.

COMMENT

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck,

M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/dbbrp/image/image.html

Insert Length: 1932 Std Error: 0.00

Seq primer: -40M13 fwd. Et from Amersham

High quality sequence stop: 477.

Location/Qualifiers

1..495

/organism="Homo sapiens"

/db_xref="taxon:9606"

FEATURES

source

```

/clone="IMAGE:1090354"
/clone_lib="NCI_CGAP_Co9"
/tissue_type="colon tumor RER+"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
RER+ colon tumor, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is not normalized. Library was
constructed by Bento Soares and M. Fatima Bonaldo
(Soares4)."
BASE COUNT      118 a   128 c   122 g   126 t       1 others
ORIGIN

Query Match      88.2%; Score 15; DB 35; Length 495;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCAGCAACTCTCT 15
|||||
Db 403 CTCAGCAACTCTCT 389

RESULT 6
LOCUS C28914 385 bp mRNA EST 06-AUG-1997
DEFINITION C28914 Rice callus CDNA Oryza sativa cDNA clone C62810_1A, mRNA
sequence.
ACCESSION C28914
NID 92312759
VERSION C28914.1 GI:2312759
KEYWORDS EST.
SOURCE Oryza sativa.
ORGANISM Oryza sativa.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza.
AUTHORS Yamamoto,K. and Sasaki,T.
TITLE Rice cDNA from callus 1997
JOURNAL Unpublished (1997)
COMMENT On Sep 12, 1996 this sequence version replaced gi:13933995.

Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai,Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abrr.affrc.go.jp.
FEATURES
source
1..385
/organism="Oryza sativa"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/map="4"
/clone="C62810_1A"
/clone_lib="Rice callus CDNA"
/tissue_type="callus"
/dev_stage="callus"
BASE COUNT      104 a   85 c   78 g   116 t       2 others
ORIGIN

Query Match      88.2%; Score 15; DB 35; Length 385;
Best Local Similarity 100.0%; Pred. No. 1.6e-02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCAGCAACTCTCT 15
|||||
Db 401 CTCAGCAACTCTCT 387

RESULT 8
LOCUS D34140/c 366 bp mRNA EST 08-AUG-1994
DEFINITION D34140 CELK041E9R Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk41e9 3', mRNA sequence.
ACCESSION D34140
NID 9525049

```

```

Qy 3 CAGCAACACTCTAT 17
|||||
Db 93 CAGCAACACTCTAT 107

RESULT 7
LOCUS AI290208/c 438 bp mRNA EST 29-JAN-1999
DEFINITION q179g10.x1 Soares.NhMpu_S1 Homo sapiens cDNA clone IMAGE:1878594
3' similar to contains Alu repetitive element.; mRNA sequence.
ACCESSION AI290208
NID G3931862
VERSION AI290208.1 GI:3931862
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 438)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Aug 21, 1998 this sequence version replaced.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 671 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 438.
FEATURES
source
1..438
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1878594"
/clone_lib="Soares.NhMpu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/notes="Organ: mixed (see below); Vector: pT7T3D-Pac
(Pharmacia) with a modified polylinker; Site.1: Not I;
Site.2: Eco RI; Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbHM, pregnant uterus
NBHPU, and fetal heart NBH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of I.M.A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."
BASE COUNT      102 a   116 c   111 g   109 t
ORIGIN

```

```

Query Match      88.2%; Score 15; DB 44; Length 438;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CTCAGCAACTCTCT 15
|||||
Db 401 CTCAGCAACTCTCT 387

RESULT 8
LOCUS D34140/c 366 bp mRNA EST 08-AUG-1994
DEFINITION D34140 CELK041E9R Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
clone yk41e9 3', mRNA sequence.
ACCESSION D34140
NID 9525049

```


VERSION	D34140.1	GI:525049
KEYWORDS	EST.	
SOURCE	Caenorhabditis elegans.	
ORGANISM	Caenorhabditis elegans	
	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.	
REFERENCE	1 (bases 1 to 366)	
AUTHORS	Kohara,Y., Mitsuiki,H., Nishigaki,A., Motohashi,T., Sugimoto,A. and Tabara,H.	
TITLE	Toward an expression map of the C.elegans genome	
JOURNAL	Unpublished (1994)	
COMMENT	On Sep 21, 1992 this sequence version replaced gi:276221.	

Contact: Yuji Kohara
 Gene Library Lab
 National Institute of Genetics
 Yata 1111, Mishima, Shizuoka 411, Japan
 Tel: 0559-75-0771
 Fax: 0559-75-6240
 Email: ykohara@dbj.nig.ac.j.
 Location/Qualifiers
 1..366
 FEATURES
 source

```

/organism="Caenorhabditis elegans"
/strain="CE1489 nim-8(e1489)"
/db_xref="taxon:6239"
/clone="yk41e9"
/clone_lib="Yuji Kohara unpublished cDNA"
/note="dev_stage=varied, sex=Hermaphrodite male,
tissue_type=whole animal"
135 a 48 c 92 g 89 t 2 others
BASE COUNT

```

```

Query Match      84.7%  Score 14.4;  DB 20;  Length 366;
Best Local Similarity 88.2%;  Pred. No. 3.3e+02;
Matches 15;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;

```

RESULT	9
N58629/c	
LOCUS	N58629
DEFINITION	235 bp mRNA EST
	28-JAN-1997
	Soares fetal liver spleen INFLS Homo sapiens cDNA clone
	IMAGE:246862 5', mRNA sequence.

ACCESSION	N58629		
NID	GL202519		
VERSION	N58629.1	GI:1202519	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 235)		
AUTHORS	Chilless, S., Dietrich, N., Dubouque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Roehling, T., Schellenberg, K., Soares, M. B., Tan, F., Thierry-Mieg, J., Trevasaki, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R., and Marra, M.		

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 5 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On Apr 14, 1993 this sequence version replaced gi:635823.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 1316 Std Error: 0.00
Seq primer: T7

```

high quality sequence stop: 1.
Location/Qualifiers
1. .235
/organism="Homo sapiens"
/db_xref="GDB:3796108"
/db_xref="taxon:9606"
/clone_image="IMAGE:246862"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="organ: Liver and Spleen; vector: p7T3D (Pharmacia)
with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
15' AACGTGAAGAAATTAAAGATCTTTTTTTTTTTT 3',
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified p7T3 vector. Library
went through one round of normalization. Library
constructed by Bencio Soares and M.Fatima Bonaldo."
61 a 40 c 43 g 79 t

```

Query Match	84.7%	Score 14.4;	DB 25;	Length 235;
Best Local Similarity	93.8%	Pred. No. 3e+02;		
Matches 15;	Conservative	0;	Mismatches 1;	Indels 0;
Gaps	0;			
QY	2	TCAGCAACACTCCTAT	17	
Db	67	TCAGCAACACTCCCAT	52	

[illegible]

W75479 ACCESSION
G1385758 NID
W75479.1 GI:1385758
EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 563)
Marrar,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie.T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Apr 14, 1993 this sequence version replaced gi:692592.

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:243298

```

Seq primer: ETPRimer
High quality sequence stop: 325.
Location/Qualifiers
1. .563
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="7"
/clone="IMAGE:391466"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCACTCTGAAGTGGAGCGCGCGAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru KO, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo. "
BASE COUNT      137 a 124 c 163 g 139 t
ORIGIN

Query Match      84.7%; Score 14.4; DB 26; Length 563;
Best Local Similarity 93.8%; Pred. No. 3.6e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 CTCAGCAACTCTCTCTA 16
|||||
Db 562 CTCAGCAACTCTCTCTA 547

RESULT 11
LOCUS      AA391390      629 bp      mRNA      EST      27-NOV-1998
DEFINITION LD10274.5prime LD Drosophila melanogaster embryo Bluescript
Drosophila melanogaster cDNA clone LD10274 5prime, mRNA sequence.
ACCESSION      AA391390
NID            92790692
VERSION        AA391390.1 GI:2790692
KEYWORDS       EST.
SOURCE         fruit fly.
ORGANISM       Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophiliidae; Drosophila.
1 (bases 1 to 629)
Harvey, D., Hong, L., Evans-Holm, M., Pendleton, J., Su, C.,
Brokstein, P., Lewis, S. and Rubin, G.M.
BDGP/HMI Drosophila EST Project
Unpublished (1997)
JOURNAL
COMMENT        On Jan 17, 1998 this sequence version replaced gi:2044375.

Contact: Harvey, D.
G. M. Rubin-Molecular and Cell Biology
University of California Berkeley
539 LSA, Berkeley, CA 94720-3200, USA
Fax: 510 643 9947
Email: http://www.fruitfly.org/EST, est@fruitfly.berkeley.edu
Plate: 102 row: G column: 2
High quality sequence stop: 445.
Location/Qualifiers
1. .629
/organism="Drosophila melanogaster"
/db_xref="BDGP_EST:BDcin009541"
/db_xref="taxon:7227"

FEATURES
source

Seq primer: ETPRimer
High quality sequence stop: 325.
Location/Qualifiers
1. .563
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="7"
/clone="IMAGE:391466"
/clone_lib="Soares mouse embryo NbME13.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCACTCTGAAGTGGAGCGCGCGAATTTTTTTTTTTTTTTTTTTT
T 3'], on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru KO, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
p7T73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo. "
BASE COUNT      137 a 124 c 163 g 139 t
ORIGIN

Query Match      84.7%; Score 14.4; DB 38; Length 629;
Best Local Similarity 93.8%; Pred. No. 3.6e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCAGCAACTCTCTCTAT 17
|||||
Db 70 TCAGCAACTCTCTAT 55

RESULT 12
LOCUS      AF038251      2374 bp      mRNA      EST      30-MAR-1998
DEFINITION AF038251 Human mRNA (Tripodis and Ragoussis) Homo sapiens cDNA
clone ntBlcon, mRNA sequence.
ACCESSION      AF038251
NID            92815881
VERSION        AF038251.1 GI:2815881
KEYWORDS       EST.
SOURCE         human.
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 2374)
Tripodis, N. and Ragoussis, J.
Generation of a transcription map in the region immediately
centromeric to human MHC across the 6p21.2-6p21.3 chromosomal
boundary
Unpublished (1997)
COMMENT        On Jan 19, 1998 this sequence version replaced gi:2045088.

Contact: Tripodis, Nikos
Division of Medical and Molecular Genetics
Guys Hospital
7th floor, Guy's Tower, London SE1 9RT, UK
Email: nikos@ki.nl.
Location/Qualifiers
1. .2374
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="6p21.3"
/clone="ntBlcon"
/clone_lib="Human mRNA (Tripodis and Ragoussis)"
BASE COUNT      495 a 584 c 628 g 667 t
ORIGIN

Query Match      84.7%; Score 14.4; DB 39; Length 2374;
Best Local Similarity 93.8%; Pred. No. 4.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TCAGCAACTCTCTCTAT 17
|||||
Db 942 TCAGCAACTCTCTCT 957

RESULT 13
LOCUS      AI040382/c      438 bp      mRNA      EST      28-AUG-1998
DEFINITION ox16c10.x1 Soares_fetal_liver脾_1NF1LS_S1 Homo sapiens cDNA

```


/db_xref="taxon:7227"
/clone="LP01318"
/clone_lib="LP Drosophila melanogaster larval-early pupal
pOT2"
/sex="male and female"
/dev_stage="larvae-pupae"
/lab_host="DH5-alpha"
/note="Organ: whole body; Vector: pOT2; Site_1: EcoRI;
Site_2: XhoI; Sized fractionated cDNAs were directly
ligated into pOT2. Plasmid cDNA library."

BASE COUNT 195 a 151 c 168 g 119 t
ORIGIN

Query Match 84.7%; Score 14.4; DB 44; Length 633;
Best Local Similarity 93.8%; Pred. No. 3.6e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TCAGCAACACTCCTAT 17
 |||||
Db 75 TCAGCAACACTACTAT 60

Search completed: September 18, 1999, 06:47:40
Job time: 3657 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 15:48:49 ; Search time 436.05 Seconds
(without alignments)
145.869 Million cell updates/sec

Title: US-09-037-472-4

Perfect score: 20
Sequence: 1 GTTAGGAATCTTCCCACTT 20

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_om.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_ph.*
- 7: gb_pl1.*
- 8: gb_pl2.*
- 9: gb_pr1.*
- 10: gb_pr2.*
- 11: gb_pr3.*
- 12: gb_ro.*
- 13: gb_st.*
- 14: gb_sts.*
- 15: gb_sy.*
- 16: gb_un.*
- 17: gb_vi.*
- 18: em_fun.*
- 19: em_htg.*
- 20: em_hum1.*
- 21: em_hum2.*
- 22: em_in.*
- 23: em_om.*
- 24: em_or.*
- 25: em_ov.*
- 26: em_pat.*
- 27: em_ph.*
- 28: em_pl.*
- 29: em_ro.*
- 30: em_sts.*
- 31: em_sy.*
- 32: em_un.*
- 33: em_vi.*
- 34: gb_htg1.*
- 35: gb_htg2.*
- 36: gb_in1.*
- 37: gb_in2.*
- 38: em_ba1.*
- 39: em_ba2.*
- 40: em_hum3.*
- 41: em_hum4.*
- 42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	20	100.0	20	5 I73228	I73228 Sequence 4

c	2	20	100.0	9721	10	HSILIB	X04500 Human gene
c	3	16.8	84.0	1895	3	AF026543	AF026543 Bos tauru
c	4	16.8	84.0	139049	11	AC004147	AC004147 Homo sapi
c	5	16.8	84.0	42042	36	CEC3484	278059 Caenorhabdi
c	6	16.8	84.0	161348	42	AC007535	AC007535 Homo sapi
c	7	16.4	82.0	139166	11	AC003029	AC003029 Homo sapi
c	8	16.4	82.0	180549	35	AC007316	AC007316 Homo sapi
c	9	16.4	82.0	34347	36	CELK08A8	U38377 Caenorhabdi
c	10	16.4	82.0	80858	42	AC007360	AC007360 Homo sapi
c	11	16	80.0	6999	8	LEU65668	U65668 Lycopersico
c	12	16	80.0	51953	8	LEU81378	U81378 Lycopersico
c	13	16	80.0	16412	9	GMITG	X93347 G.gorilla m
c	14	16	80.0	16364	9	GORMTC	D38114 Gorilla m
c	15	16	80.0	164296	34	HSAC000380	AC000380 Homo sapi
c	16	16	80.0	158772	34	HSJ169P22	AL049588 Homo sapi
c	17	16	80.0	250111	35	AC007622	AC007622 Homo sapi
c	18	16	80.0	174652	42	AC007540	AC007540 Homo sapi
c	19	15.8	79.0	14704	2	AE001112	AE001112 Archaeogl
c	20	15.8	79.0	2150	4	AF071349	AF071249 Danio rer
c	21	15.8	79.0	180	4	DRHOXA9	Y07696 D.ferio HOX
c	22	15.8	79.0	3186	5	I22487	I22487 Sequence 23
c	23	15.8	79.0	3660	7	ATU19382	U19382 Arabidopsi
c	24	15.8	79.0	41173	7	SPBC16A3	AL021748 S.pombe c
c	25	15.8	79.0	37777	7	SPBC215	AL033534 S.pombe c
c	26	15.8	79.0	60000	8	AF049236	AF049236 Arabidops
c	27	15.8	79.0	93959	9	HS106C24	283313 Human DNA s
c	28	15.8	79.0	132790	9	HS167A14	294721 Human DNA s
c	29	15.8	79.0	121131	9	HS345B16	AL008633 Human DNA
c	30	15.8	79.0	156432	9	HS581F7	AL022164 Human DNA
c	31	15.8	79.0	3787	9	HUMCAMPB	L12686 Homo sapien
c	32	15.8	79.0	3871	9	HUMPDE2A	M97515 Human rolip
c	33	15.8	79.0	3186	9	HUMPDEB	L20966 Human phosp
c	34	15.8	79.0	4068	9	HUMPDEG	L20971 Human phosp
c	35	15.8	79.0	116370	10	HS569M23	AL031666 Human DNA
c	36	15.8	79.0	147419	10	HS738P11	AL031736 Human DNA
c	37	15.8	79.0	176552	11	AC004617	AC004617 Homo sapi
c	38	15.8	79.0	161453	11	AC004993	AC004993 Homo sapi
c	39	15.8	79.0	103606	11	AC005004	AC005004 Homo sapi
c	40	15.8	79.0	186986	11	AC006210	AC006210 Homo sapi
c	41	15.8	79.0	3932	11	HSITGBF04	U66532 Human beta4
c	42	15.8	79.0	153460	11	HSU52111	U52111 Homo sapien
c	43	15.8	79.0	251364	12	MMAE000664	AE000664 Mus muscu
c	44	15.8	79.0	2834	12	MUSCT16B01	M60352 Mouse testu
c	45	15.8	79.0	32756	34	AC006608	AC006608 Caenorhab

ALIGNMENTS

RESULT	1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
I73228	I73228	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
LOCUS	I73228	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
DEFINITION	I73228	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
ACCESSION	I73228	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
NID	g3009367	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
VERSION	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
KEYWORDS	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
SOURCE	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
ORGANISM	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
REFERENCE	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
AUTHORS	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
TITLE	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
JOURNAL	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
FEATURES	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997
source	I73228.1	Sequence	4	from patent	US 5686246.	DNA	PAT	23-DEC-1997

BASE COUNT 4 a 5 c 3 g 8 t
ORIGIN
Query Match 100.0% Score 20; DB 5; Length 20;
Best Local Similarity 100.0% Pred. No. 2.4;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTCCCACTT 20
 Db 1 GTTTAGGAATCTCCCACTT 20

RESULT 2
 HSIL1B/C
 LOCUS 9721 bp DNA PFI 26-JUN-1997
 DEFINITION Human gene for prointerleukin 1 beta.
 ACCESSION X04500
 NID 933788
 VERSION X04500.1 GI:33788
 KEYWORDS interleukin 1 beta.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
 Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE 1 (bases 1 to 9721)
 AUTHORS Clark,B.D., Collins,K.L., Gandy,M.S., Webb,A.C. and Auron,P.E.
 TITLE Genomic sequence for human prointerleukin 1 beta: possible
 evolution from a reverse transcribed prointerleukin 1 alpha gene
 JOURNAL Nucleic Acids Res. 14 (20), 7897-7914 (1986)
 MEDLINE 87040762
 REMARK Erratum:[published erratum appears in Nucleic Acids Res 1987 Jan
 26;15(2):868]]
 COMMENT Data kindly reviewed (13-MAY-1988) by Clark B.D.
 FEATURES
 source 1..9721 Location/Qualifiers
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="2"
 /clone_lib="leukocyte DNA library"
 /map="q13-q24"
 misc_feature 669..960
 /note="Alu repeat"
 misc_feature 1576..1581
 /note="TATA-box like sequence"
 CAAT_signal 1809..1816
 CAAT_signal 1859..1866
 TATA_signal 1903..1909
 prim_transcript 1934..8953
 exon 1934..2005
 /number=1
 misc_feature complement(1936..1943)
 /note="pot. viral enhancer core sequence"
 intron 2006..2465
 /number=1
 repeat_region 2039..2055
 /note="direct repeat 2"
 misc_feature 2291..2297
 /note="pot. viral enhancer core sequence"
 misc_feature 2458..2465
 /note="pot. viral enhancer core sequence"
 exon 2466..2527
 /number=2
 CDS join(2481..2527,3092..3143,5125..5326,5874..6038,
 7275..7405,8127..8339)
 /codon_start=1
 /product="prointerleukin-1-beta"
 /protein_id="CAA28185.1"
 /db_xref="PID:g312408"
 /db_xref="GI:312408"
 /translation="MAEVPDLASEMAYISGNEDDLFEADGPKQMKCSFQDLDCPL
 DGGIQLRSDHHYSKGFQAAASVYVAMDRLKRLVPCPTQFQNDLSTFFPFIFEEP
 IFDWDNEAYVHDAPVRSNLCTLRDSQKSLVMSGPKVELKALHQQGDMEOQVFSM
 SFVQGEESNDKIPVALGLKELKLYLSVLDKDKPTLOLESVDPKNYPKKMKRFFVN
 KIEINNKLEFESAQFPNWIYSTQAEINMPVFLGGTKGGODITDTMQFVSS"
 2528..3091
 /number=2
 intron complement(2714..2721)
 misc_feature

misc_feature /note="pot.viral enhancer core sequence"
 2858..2865
 /note="pot.viral enhancer core sequence"
 3092..3143
 /number=3
 intron 3144..5124
 /number=3
 misc_feature 4266..4273
 /note="pot.viral enhancer core sequence"
 4274..4279
 /note="pot.viral enhancer core sequence"
 4659..4988
 /note="Alu repeat"
 5125..5326
 /number=4
 intron 5327..5873
 /number=4
 misc_feature 5489..5497
 /note="pot.viral enhancer core sequence"
 5874..6038
 /number=5
 intron 6039..7274
 /number=5
 repeat_unit 6205..6212
 /note="inverted repeat A"
 6213..6220
 /note="inverted repeat A"
 complement(6910..6915)
 /note="pot.viral enhancer core sequence"
 complement(6916..6924)
 /note="pot.viral enhancer core sequence"
 complement(7247..7253)
 /note="pot.viral enhancer core sequence"
 7275..7405
 /number=6
 intron 7406..8126
 /number=6
 misc_feature 7419..7426
 /note="pot. viral enhancer core sequence"
 8127..8953
 /number=7
 polyA_signal 8925..8930
 /note="pot polyA signal"
 polyA_site 8953
 misc_feature 9331..9721
 /note="Alu repeat"
 BASE COUNT 2661 a 2328 c 2122 g 2608 t 2 others
 ORIGIN

Query Match 100.0%; Score 20; DB 10; Length 9721;
 Best Local Simlity 100.0%; Pred. No. 1;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTCCCACTT 20
 Db 1536 GTTTAGGAATCTCCCACTT 1517
 |||||

RESULT 3
 AF026543/c
 LOCUS 1895 bp DNA MAM 30-JUN-1998
 DEFINITION Bos taurus interleukin-1 beta (IL-1 beta) gene, 5' flanking region
 and partial cds.
 ACCESSION AF026543
 NID g3258624
 VERSION AF026543.1 GI:3258624
 KEYWORDS Bos taurus.
 SOURCE Bos taurus.
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.

REFERENCE	1 (bases 1 to 1895)
AUTHORS	Tseng,Y.-H. and Schuler,L.A.
TITLE	Transcriptional regulation of interleukin-1beta gene by interleukin-1beta itself is mediated in part by Oct-1 in thymic stromal cells
JOURNAL	J. Biol. Chem. 273 (20), 12633-12641 (1998)
MEDLINE	98241640
REFERENCE	2 (bases 1 to 1895)
AUTHORS	Tseng,Y.-H. and Schuler,L.A.
TITLE	Direct Submission
JOURNAL	Submitted (24-SEP-1997) Comparative Biosciences, University of Wisconsin-Madison, 2015 Linden Drive West, Madison, WI 53706, USA
FEATURES	Location/Qualifiers
source	1..1895 /organism="Bos taurus" /db_xref="taxon:9913" join(1211..1255,1875..>1895) /gene="IL-1 beta" 1211..>1895 /gene="IL-1 beta" 1890..>1895 /gene="IL-1 beta" /note="cytokine" /codon_start=1 /product="interleukin-1 beta" /protein_id="AAC24693.1" /db_xref="PID:g3258625" /db_xref="GI:3258625" /translation="MA"
BASE COUNT	511 a 472 c 363 g 549 t
ORIGIN	
Query Match	84.0% Score 16.8 DB 3 Length 1895;
Best Local Similarity	90.0% Pred.No.59;
Matches 18:	Conservative 0 Mismatches 2 Indels 0 Gaps 0
Qy 1	GTTTAGGAATCTTCCCACTT 20
Db 808	GTTTAGGAACCTTCCTACTT 789
RESULT 4	
AC004147/c	
LOCUS	AC004147 139049 bp DNA PRI 28-MAR-1998
DEFINITION	Homo sapiens chromosome 17, clone CITI304117, complete sequence.
ACCESSION	AC004147
NID	g2995605
VERSION	AC004147.1 GI:2995605
KEYWORDS	HTG.
SOURCE	human.
ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
1 (bases 1 to 139049)	
Birren,B., Fasnman,K., McKernan,K., Nusbaum,C., Richardson,P. and Lander,E.	
Homo sapiens chromosome 17, clone CITI304117	
Unpublished	
2 (bases 1 to 139049)	
Birren,B., Fasnman,K., McKernan,K., Nusbaum,C., Richardson,P., Lander,E., Allen,N., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Boutwell,C., Byrne,S., Cantu,C., Castle,A., Cerny,J., Cooke,P., Daly,W.J., Depayre,E., Devon,K., Dewar,K., Donelan,L., DuRette,B., Etemadi,S., Ferreira,P., Forrest,C., Funke,R., Gage,D., Gardyna,S., Gensheimer,S., Gerlaigry,K., Gilmarin,T., Gray,D., Hagos,B., Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Linton,L., MacKenzie,J., Marquis,N., McEwan,P., McGurk,A., Meldrim,J., Molla,M., Morris,W., Morrow,J., Nachman,A., Naylor,J., O'Connor,T., Pavlin,B., Peterson,K., Ranganath,S., Riley,R., Roberts,D., Rollins,G., Rossello,R., Roy,A., Shiyam,R., Soohoo,S., Stange-Thomann,N., Stillwell,J., Stone,C., Strickland,C., Sydney,K., Tang,L., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y.,	

repeat_region	/rpt_family="(CA)n" complement(15748..15780)
repeat_region	/rpt_family="(CA)n" 18631..18719
repeat_region	/rpt_family="MIR" complement(19314..19340)
repeat_region	/rpt_family="AT-rich" 19383..19529
repeat_region	/rpt_family="MIR" complement(20229..20513)
repeat_region	/rpt_family="AluSg" complement(20804..21671)
repeat_region	/rpt_family="L1MA8" 22031..22246
repeat_region	/rpt_family="AluJb" 22247..22279
repeat_region	/rpt_family="(GAAAA)n" 22302..22324
repeat_region	/rpt_family="AT-rich" 24057..24121
repeat_region	/rpt_family="MIR" complement(24283..24316)
repeat_region	/rpt_family="(TAAA)n" 24692..24977
repeat_region	/rpt_family="L1RI6A" 26655..26784
repeat_region	/rpt_family="MIR" complement(27085..27296)
repeat_region	/rpt_family="MIR" complement(27385..27537)
repeat_region	/rpt_family="MSTC" complement(27752..28040)
repeat_region	/rpt_family="MSTC" complement(28043..28164)
repeat_region	/rpt_family="MIR" complement(29708..29846)
repeat_region	/rpt_family="LINE2" complement(29966..30050)
repeat_region	/rpt_family="L1MC4" complement(30389..31942)
repeat_region	/rpt_family="L1MC4" complement(33029..33197)
repeat_region	/rpt_family="MIR" 33376..33733
repeat_region	/rpt_family="MLT1A1" complement(34560..35178)
repeat_region	/rpt_family="MER4D" complement(35177..35407)
repeat_region	/rpt_family="MER4D" 35547..35578
repeat_region	/rpt_family="AT-rich" complement(35749..35909)
repeat_region	/rpt_family="MER5B" complement(37344..37553)
repeat_region	/rpt_family="MER20" 38656..39153
repeat_region	/rpt_family="L1RI2" complement(39154..39253)
repeat_region	/rpt_family="MIR" complement(40534..40644)
repeat_region	/rpt_family="LINE2" 40801..40883
repeat_region	/rpt_family="LINE2" 42080..42224
repeat_region	/rpt_family="MIR" 42430..42528
repeat_region	/rpt_family="MER5B" complement(42774..42839)
repeat_region	/rpt_family="MIR" complement(43354..43413)
repeat_region	/rpt_family="MIR" 44462..44552
repeat_region	/rpt_family="MER5A"
repeat_region	45602..45680
repeat_region	/rpt_family="MIR" 46417..46803
repeat_region	/rpt_family="MLT1A2" complement(47106..47146)
repeat_region	/rpt_family="MIR" 47153..47250
repeat_region	/rpt_family="MIR" complement(47251..47551)
repeat_region	/rpt_family="AluSq" complement(47626..47771)
repeat_region	/rpt_family="MIR" 49002..49210
repeat_region	/rpt_family="MIR" complement(49715..49904)
repeat_region	/rpt_family="LINE2" complement(50021..50308)
repeat_region	/rpt_family="MLT1C" 50312..50614
repeat_region	/rpt_family="AluSx" complement(50630..50798)
repeat_region	/rpt_family="MLT1C" complement(50809..50928)
repeat_region	/rpt_family="MST-INTERNAL" complement(51018..51119)
Query Match 84.0%; Score 16.8; DB 11; Length 139049;	
Best Local Similarity 90.0%; Pred. NO. 32;	
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;	
Qy 1	GTTTAGGAATCTTCCCACTT 20
Db 98475	GTTTAGGTTTCTTCCCACTT 98456
RESULT 5	
LOCUS CEC34B4/c	
DEFINITION CEC34B4 42042 bp DNA INV 23-NOV-1998	
ACCESSION 278059	
NID g1487886	
VERSION 278059.1 GI:1487886	
KEYWORDS HTG.	
SOURCE Caenorhabditis elegans.	
ORGANISM Caenorhabditis elegans	
REFERENCE 1 (bases 1 to 42042)	
AUTHORS Smye,R.	
TITLE Direct Submission	
JOURNAL Submitted (09-AUG-1996) Louis, MO 63110, USA. E-mail: jes@sanger.ac.uk or rw@nematode.wustl.edu	
REFERENCE 2 (bases 1 to 42042)	
AUTHORS Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latreille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Shownkeen,R., Smaildon,N., Smith,A., Sonnhammer,E., Staaden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohldman,P.	
TITLE 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans	
JOURNAL Nature 368 (6466), 32-38 (1994)	
MEDLINE 94150718	
COMMENT Coding sequences below are predicted from computer analysis, using predictions from Genefinder (P. Green, U. Washington), and other available information.	
For a graphical representation of this sequence and its analysis see: -	

http://webcace.sanger.ac.uk/cgi-bin/display7db-wormaceclass=Sequence subject-C34B4

Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.

IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.

This sequence is the entire insert of clone C34B4. The true right end of clone F4ZE8 is at 18123 in this sequence. The start of this sequence (1..101) overlaps with the end of sequence Z78014. The end of this sequence (41942..42042) overlaps with the start of sequence Z83242.

```
FEATURES
    source
        1..42042
            /organism="Caenorhabditis elegans"
            /db_xref="taxon:6239"
            /chromosome="v"
            /clone="C34B4"
            13357..28308
            /gene="C34B4.2"
            join(13357..13641,15622..15705,16741..16899,17560..17701,
            17896..18021,18076..18192,18240..18352,18402..18500,
            19332..19380,19982..20215,21671..21737,21913..21989,
            22062..22194,22338..22435,22520..22697,22743..22864,
            22910..22989,23769..23831,24153..24285,24335..24553,
            24599..24782,24937..25070,25115..25275,25777..25903,
            27238..27347,27585..27765,27825..27977,28031..28114,
            28159..28308)
            /gene="C34B4.2"
            /note="predicted using GeneFinder; similar to Activin
            types I and II receptor domain"
            /codon_start=1
            /protein_id="CAB01486.1"
            /db_xref="PID:e1344485"
            /db_xref="PID:g3874640"
            /db_xref="GI:3874640"
            /db_xref="SPTREMBL:Q18419"
            /translation="MREYILRLRLRGSTDKDVLIRIDRLSKSRSTSAISLVLEGCL
            RANRPNRCSLFTSFFSSLSALFVCDVITGARTSFSKSTICKTPTPEAYVFNQD
            RNSGKTDGVLQGLAAKCKMSHRNSOYFREIVAFSGFAGNRLCLRHVDILSIPO
            KRLVTOHALPDEAQTQTSNLMWAFDMDEHSEMMLSANSPDDDLQWQFDQ
            IIKENDLSAASPSSLSDFDDACTSSFASLSPTFATSSGTSSEGTSLPQFDDMM
            SMFPDPTPTVDDRMAPPATYKKNVRRPRECMVCLFRTLLTSYRALPANA
            IFEFTEETMOQSTAGSARHYRDSIRHCLMTSAVFVRVLTNPDKDFLVMNEQSWTIV
            HPDCMACAEHVRSGSAHRCPPOASTPAYAVRMEETKAASSGSTACPIRATGCV
            QSASSSCSTSSSSASPAPSVNREFDPIRKVTYGTMGNGORVSKAFSSCTCGD
            FCVARRSGAVSLIEDRLCITGPKMPYFCTLLWDELICYNCTDCNSNKLKYSNT
            VLPVIECKQVHNNRYKLVTCNKLRIISYTKNRIGPGREDHEBKIOCSILGESADFYI
            DTNSFDEMTARNFYETACYNVSHQHFYFCRCQTNCSNPSPIPYPLSPPTVSC
            FTSGFEAYVPTKYSTNTWHVONYKKNESYVDENSTCHGDMGFIKAAVADAGEY
            DFVYVYKGCISTNHNHNSIOHNRLEVRKNMKTNDENPVLQFTRCTRVSSGSKLL
            IHAHWIQLICTDQCNITISLLRGNITILPVIOCKALIQTVYTAIPSNKCRIRIISYIKH
            DRYELDFENDKVDLQEFISGAERGSGLVIDAHYIWPFEKMLRNFPSFGCCYNTDMS
            PRHYIYVCRCEANCTPESEALGPISPPVVKCHTSGFAYEMNRYFKELETGLDNL
            YKQNYMTNIEYDGKFPSECOGHFCVIVPNKYIIDGEYVDIYKQISANQOGEOK
            LSVQKYNVEITYYICHTDFCNIDTSLDAANASTISDNGADGTFKDFCFQVILFR
            RSRHNDNCDRLVLIAQRTLAYSTVWGRHLHSTCFPMKTFWFLLOITLAFILRS
            TNLNVNPNATLFSVAEQPIYIWMVVLPSLTFLFGQDNCKMYRFNDRQKFTL
            TCFVNTKWKIEIPRIQIIEITIOIPIIILYIALVIKIKMKYTSRKTETI
            IILKQAFIFLMFOIYNIIVLYGKLTAKMTVTAFYILKRAITHAIEFAGATTPCFVFT
            SKEIKLVNKNITAGSGNSVLRNRIRASVPRNS"
            complement(29973..33985)
            /gene="C34B4.1"
            complement(join(29973..30197,30248..30446,30680..31040,
            31165..31364,31739..32462,32683..32875,32920..33606,
            33764..33873,33937..33985))
            /gene="C34B4.1"
            /note="similar to PH (pleckstrin homology) domain; cDNA
            EST EMBL:D33476 comes from this gene; cDNA EST EMBL:T00122
```

comes from this gene: cDNA EST EMBL:D36335 comes from this gene: cDNA EST EMBL:D67933 comes from this gene: cDNA EST yk260h6.3 comes from this gene: cDNA EST yk260h6.5 comes from this gene: cDNA EST yk243d7.3 comes from this gene: cDNA EST yk243d7.5 comes from this gene: cDNA EST yk224q8.3 comes from this gene: cDNA EST yk209h12.3 comes from this gene: cDNA EST yk209h12.5 comes from this gene"

```
/codon_start=1
/protein_id="CAB01487.1"
/db_xref="PID:e1344486"
/db_xref="PID:g3874641"
/db_xref="GI:3874641"
/db_xref="SPTREMBL:Q18420"
/translation="MSVSISETPRVQKAAKATMLKNTSQSGSTSESSPCEDEAKPMP
LPRIHPATERDSLSNAEDYEDIEESFERHEEQVNRNEKYVDYVNLHEFVAIK
DKVYSEIVKHSRTEKEYLPALPSHQPRQWENRILQAAKCLSVIDGADSGTGTGP
SVSPYHSDVSRISSNNRTPSGSKSLSRNSPVPVSHSHSYEDGAYALRDLCGL
SDYILPPDASSRCSGTPTSRSLIPSRTEKSGYWTQTDSTRISKLRRVYVFKNG
QISFTRKNNRDEEPASKISIDIRSVYKLEQQAAYAFOLITSDKNFMTESERTT
HDMVILSAAIKATTLREMASTVPIDASISGWLTVKCGLSKKVFAALVNOKLMFFK
NSDLVPNGFLCLOQAISERKNGTEYSGSDEQLETTKEHPNQRKNNDSLCVOIAN
EDPVYILRTSDEKELWYLYKSASGTAALCGTPEILVQRMMAENVANGMIFHNEDR
LFTWISDPSLWKDLLEFASGEIPKDTMTSVDSRKKTLLEIARACQLFVSLMRAQA
TOYHDLAQNILSTAVQOYELRNESEDETRMAIFCESALDRCLRVGRGROGSRLEVTS
VLTRDYTKPHSISVYKLPNSEYOIVFEDGKOKLCLDALSTWETSRDAQRGVISED
AAALSRLRRHWGHLALTETPIERQFLVNRASEEIVNGRIPUSNOLCESLAALYQMV
FGDNQNSDDQGFESISORFYPSKMLDVACIKSLRQINSSEILGENSECVRVIL
QVILGKPLFGDLHEASMTDNERIYIALNDHSVSLDRRHWFIVIRTPYSSLSFTFG
QFOQDFMLTIIRPLTPGSHPEAKRLETFSMQKNEIEQLTLHLAEYIRQCQIVNKRVS
K"
BASE COUNT 14077 a 7115 c 7156 g 13694 t
ORIGIN
```

```
Query Match 84.0%; Score 16.8; DB 36; Length 42042;
Best Local Similarity 90.0%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTAGGAATCTTCCCACTT 20
      ||||| ||||| |||||
Db 20441 GTTAGGAGCTTCCCACTT 20422

RESULT 6
AC007535/c 161348 bp DNA 29-MAY-1999
LOCUS Homo sapiens 12p13 BAC RPC111-154121 (Roswell Park Cancer Institute
DEFINITION Human BAC Library) complete sequence.
ACCESSION AC007535
NID 94914349
VERSION AC007535.2 GI:4914349
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 161348)
Muzny,D., Arenson,A.D., Bouck,J., Bunac,C., Chen,Z., Ding,Y.,
Dugan,S., Durbin,J., Forcum,J., Garcia,C., Gorrell,J.H.,
Gorrell,L.L., Hernandez,J., Issar,A., Jackson,L., Kneitz,S.,
Kondejewski,N., Lau,S., Leal,B., Lee,E., Lichtarge,O., Liu,W.,
Logan,O., Lu,J., Marondel,I., Martinez,C., Merscher,S., Miller,A.,
Montgomery,K., Oswal,G., Pampell,L.R., Parish,B.J., Perez,L.,
Rashid,N.D., Rives,C., Scherer,S.E., Shen,H., Shim,C., Simon,M.,
Vo,Q., Williamson,A., Worley,K.C., Xhang,A.M., Yang,R., Yu,W.,
Zhou,X., Kucherlapati,R., Nelson,D. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 161348)
Worley,K.C.
Direct Submission
Submitted (13-MAY-1999) Molecular and Human Genetics, Baylor
```



```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC111-154121"
/chromosome="12p13"
243..381
/rpt_family="L2"
456..530
/rpt_family="MIR"
844..923
/rpt_family="MIR"
complement(4368..4503)
/rpt_family="MIR"
7912..8298
/rpt_family="THE1C"
complement(8498..8642)
/rpt_family="MLT1F"
9314..9606
/rpt_family="L2"
10218..10515
/rpt_family="AluSc"
complement(1494..14543)
/rpt_family="(TA)n"
14626..14720
/rpt_family="MIR"
14797..14843
/rpt_family="L2"
15478..15607

Query Match      84.0%; Score 16.8; DB 42; Length 161348;
Best Local Similarity 90.0%; Pred. NO. 31;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTTAGGATCTTCCACTT 20
||||| ||||| ||||| |||
DB 98301 GTTTAAGAATCTCCCTCTT 98282

RESULT 7
AC003029 AC003029 139166 bp DNA PRI 17-SEP-1998
LOCUS Homo sapiens Chromosome 12q24 PAC RPC13-462E2 (Roswell Park Cancer
DEFINITION Institute Human PAC library) complete sequence.
ACCESSION AC003029
NID 93366554
VERSION AC003029.1 GI:3366554
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; C-aniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 139166)
AUTHORS Muzny,D., Arenson,A.D., Adams,C., Bunac,C., Carvelli,K., Chang,J.,
Chacko,J., Chen,J., Ding,Y., Dugan,S., Durbin,J., Forcum,J.,
Ganesh,R., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M.,
Hernandez,J., Jackson,L., Jin,S., Kampal,R., Karpthy,S., Kovar,C.,
Lau,S., Leal,B., Lee,E., Li,Y., Lichtarge,O., Liu,W., Logan,O.,
Lu,J., Ly,T., Marondel,I., Martinez,C., Merscher,S., Montgomery,K.,
Oswal,G., Perez,L., Rashid,N.D., Renault,B., Rowland,K., Savage,L.,
Scherer,S.E., Shen,H., Simon,M., Stovall,K., Timms,K.M., Todd,J.,
Vo.O., Williamson,A., Worley,K.C., Yu,W., Kucherlapati,R.,
Nelson,D. and Gibbs,R.A.
Direct Submission
2 (bases 1 to 139166)
Worley,K.C.
Direct Submission
Submitted (27-OCT-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 139166)
Worley,K.C.
Direct Submission
Submitted (31-JUN-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

TITLE
JOURNAL
AUTHORS
Worley,K.C.
Submitted (16-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
5 (bases 1 to 139166)
Worley,K.C.
Direct Submission
Submitted (17-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jul 31, 1998 this sequence version replaced gi:2909690.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
FEATURES
source
1..139166
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC13-462E2"
/chromosome="12q24"
1..226
/notes="Overlaps with bases 1-226 in AC002996."
/notes="Region: Overlap with AC002996"
join(88..201,843..945,1380..1564,6860..7024,8688..8845,
12641..12755,18325..18739)
/gene="Human nucleus-encoded mitochondrial aldehyde
dehydrogenase(ALDH)gene"
/notes="Unigene cluster Hs.74630, K03001, X05409"
530..696
/rpt_family="AluJb"
repeat_region
1648..1732
/rpt_family="MIR"
repeat_region
complement(2001..2165)
/rpt_family="FRAM"
repeat_region
complement(2260..2370)
/rpt_family="L1MC3"
repeat_region
complement(2374..2612)
/rpt_family="AluSc"
repeat_region
complement(2631..2799)
/rpt_family="L1MC2"
repeat_region
complement(2800..3099)
/rpt_family="AluSc"
repeat_region
complement(3100..3220)
/rpt_family="L1MC2"
repeat_region
complement(3291..3588)
/rpt_family="AluJo"
repeat_region
complement(3608..4064)
/rpt_family="L1MB7"
repeat_region
complement(4069..4364)
/rpt_family="AluSc"
repeat_region
complement(4374..4672)
/rpt_family="AluSg"
repeat_region
4761..4893
/rpt_family="FLAM_C"
repeat_region
4904..5116
/rpt_family="AluSg/x"
repeat_region
5403..5684
/rpt_family="AluJb"

```

```
repeat_region complement(5744..6014)
repeat_region /rpt_family="AluSg"
6185..6402
repeat_region /rpt_family="AluSg"/%
repeat_region complement(7325..7487)
repeat_region /rpt_family="FRAM"
7507..7790
repeat_region /rpt_family="AluSx"
repeat_region complement(7800..7943)
repeat_region /rpt_family="AluSg"
repeat_region complement(7947..8248)
repeat_region /rpt_family="AluSx"
repeat_region complement(8255..8389)
repeat_region /rpt_family="FLAM_C"
repeat_region complement(9202..9433)
repeat_region /rpt_family="AluJo"
9452..9785
repeat_region /rpt_family="L1MB6"
repeat_region complement(9788..10134)
repeat_region /rpt_family="THE1B"
10141..10197
repeat_region /rpt_family="L1ME1"
10201..10493
repeat_region /rpt_family="AluSx"
10495..10754
repeat_region /rpt_family="L1MB6"
10755..10916
repeat_region /rpt_family="AluSc"
10918..11213
repeat_region /rpt_family="AluSg"
11214..11379
repeat_region /rpt_family="AluSc"
11381..11681
repeat_region /rpt_family="AluY"
11693..11988
repeat_region /rpt_family="AluSg"
11996..12094
repeat_region /rpt_family="(TAAAA)n"
12100..12229
repeat_region /rpt_family="AluJo"
12241..12529
repeat_region /rpt_family="AluSx"
complement(12530..12557)
repeat_region /rpt_family="AT-rich"
12631..12765
STS /standard_name="DWC10/DWC11"
13248..13548
repeat_region /rpt_family="AluSx"
repeat_region complement(13595..13622)
repeat_region /rpt_family="AT-rich"
13999..14103
repeat_region /rpt_family="AluJo"
complement(14148..14404)
repeat_region /rpt_family="AluDb"
14525..14811
repeat_region /rpt_family="AluSc"
14814..15118
repeat_region /rpt_family="AluY"
15472..15768
repeat_region /rpt_family="AluSx"
15767..16064
repeat_region /rpt_family="AluY"
complement(16333..16541)
repeat_region /rpt_family="L2"
repeat_region complement(16658..16956)
repeat_region /rpt_family="AluSx"
16992..17186
repeat_region /rpt_family="L1MB8"
17205..17506
repeat_region /rpt_family="AluSx"
complement(17550..17928)
repeat_region /rpt_family="L2"

17954..18213
repeat_region /rpt_family="AluJo"
18511..18642
STS /gene="Human nucleus-encoded mitochondrial aldehyde
dehydrogenase(ALDH)gene"
/standard_name="SHGC-11012"
/note="D12S1884, RH8637, STS17970, GDB:674606,
GDB:675353, Chr. 12pter-13qter between D12S84 and D12S369,
Homo Sapiens"
/db_xref="dbSTS:17970"
18807..19105
repeat_region /rpt_family="AluSx"
complement(19388..19650)
repeat_region /rpt_family="AluJo"
19837..20151
repeat_region /rpt_family="AluSp"
complement(20254..20555)
repeat_region /rpt_family="AluSx"
complement(20559..21050)
repeat_region /rpt_family="L1RL3"

Query Match 82.0%; Score 16.4; DB 11; Length 139166;
Best Local Similarity 94.4%; Pred. No. 52;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTAGGAATCTTCCCACT 19
||||||| |||||||
Db 85632 TTTAGGAAGCTTCCCACT 85649

RESULT 8
AC007316 LOCUS AC007316 180549 bp DNA HTG 05-JUN-1999
DEFINITION Homo sapiens clone NH0356B17, WORKING DRAFT SEQUENCE, 4 unordered
pieces
ACCESSION AC007316
NID 95001505
VERSION AC007316.2 GI:5001505
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 180549)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 180549)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (16-APR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Jun 5, 1999 this sequence version replaced gi:4586090.
* NOTE: this is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 36196: contig of 36196 bp in length
* 36197 36214: gap of unknown length
* 36215 74570: contig of 38356 bp in length
* 74571 74588: gap of unknown length
* 74589 124355: contig of 49767 bp in length
* 124356 124373: gap of unknown length
* 124374 180549: contig of 56176 bp in length.
Location/Qualifiers
1..180549
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source
```

BASE COUNT	53903 a	36172 c	35964 g	54456 t	54 others
ORIGIN	/clone="NH0356B17"				
Query Match	82.0%;	Score 16.4;	DB 35;	Length 180549;	
Best Local Similarity	94.4%;	Pred. No. 50;			
Matches 17; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;	
QY	2 TTTAGGAATCTTCCCACT 19				
Db	4556 TTTAGGAATCTTCCCACT 4573				
RESULT	9				
LOCUS	CELK08A8 34347 bp DNA INV 19-OCT-1995				
DEFINITION	Caenorhabditis elegans cosmid K08A8.				
ACCESSION	U38377				
NID	g1022968				
KEYWORDS	U38377.1 GI:1022968				
SOURCE	.				
ORGANISM	Caenorhabditis elegans strain=Bristol N2. Caenorhabditis elegans Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis. 1 (bases 1 to 34347) Wilson,R., Ainscough,R., Anderson,K., Baynes,C., Berks,M., Bonfield,J., Burton,J., Connell,M., Copsey,T., Cooper,J., Coulson,A., Craxton,M., Dear,S., Du,Z., Durbin,R., Favello,A., Fulton,L., Gardner,A., Green,P., Hawkins,T., Hillier,L., Jier,M., Johnston,L., Jones,M., Kershaw,J., Kirsten,J., Laister,N., Latrille,P., Lightning,J., Lloyd,C., McMurray,A., Mortimore,B., O'Callaghan,M., Parsons,J., Percy,C., Rifken,L., Roopra,A., Saunders,D., Showkhen,R., Smaildon,N., Smith,A., Sonnenhammer,E., Staden,R., Sulston,J., Thierry-Mieg,J., Thomas,K., Vaudin,M., Vaughan,K., Waterston,R., Watson,A., Weinstock,L., Wilkinson-Sproat,J. and Wohlman,P. 2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans 368 (6466), 32-38 (1994)				
JOURNAL	Nature 368 (6466), 32-38 (1994)				
MEDLINE	94150718				
REFERENCE	2 (bases 1 to 34347)				
AUTHORS	Pauley,A.				
TITLE	The sequence of C. elegans cosmid K08A8				
JOURNAL	Unpublished (1995)				
REFERENCE	3 (bases 1 to 34347)				
AUTHORS	Waterston,R.				
TITLE	Direct Submission				
JOURNAL	Submitted (11-OCT-1995)				
COMMENT	Submitted by: Genome Sequencing Center Department of Genetics, Washington University, St. Louis, MO 63110, USA, and Sanger Centre, Hinxton Hall Cambridge CB10 1RQ, England e-mail: rwenematode.wustl.edu and jes@sanger.ac.uk				

NEIGHBORING COSMID INFORMATION:

5' cosmid is C01C10, 200bp overlap; 3' cosmid is F48E3, 400bp overlap. Actual start is at base position 197 of CELK08A8; actual end is at base position 4741 of CELF48E3.

NOTES:

Coding sequences below are predicted from computer analysis, using the program Genefinder (P. Green and L. Hillier, ms in preparation).

FEATURES

Location/Qualifiers

Query Match

82.0%; Score 16.4; DB 36; Length 34347;

	BASE COUNT
	ORIGIN

1. .34347
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/map="x"
/strain="Bristol N2"
complement(10814. .12280)
/gene="K08A8.2"
complement(join(10814. .11194,11247. .11351,11423. .11548,
11990. .12053,12132. .12280))
/gene="K08A8.2"
/note="similar to mammalian sex-determining protein SRY"
/codon_start=1
/evidence=not_experimental
/protein_id="AA79747.1"
/db_xref="PID:g1022970"
/db_xref="GI:1022970"
/translation="MODYSGWGFHMPPTSSNTLLPQSPDLSNGPDSPTSGDKG
KNKDRGKPMNAFVMSQGRKMALEKNKNSISKRGLTGTNKMVLEQEKRPFD
EAKLRRAIMKSHPDYKYRPRKTKSINKNGASIPFGLNLDTKTSPYLTITNNATN
QYIDRFAPYPTTIVDOIPLSTVPVHSVTDNNSPQSPQSPNSTFAGSYLTPK
SESSVGSOSTVGVDSOSFQYRITDIKQMPPISEIETHAQNIQNALSQSHVIS"
complement(11118. .27091)
/gene="K08A8.3"
complement(join(11118. .18288,18995. .19999,20892. .21092,
22889. .23029,23078. .23210,24517. .25085,26088. .26237,
26429. .26608,26665. .26809,26928. .27091))
/gene="K08A8.3"
/note="weak similarity to M. musculus ubiquitin
carboxyl-terminal hydrolase (SP:UBP_MOUSE, P35123);
similar to S. pombe double-strand-break repair protein,
RAD21 (SP:RA21_SCHPO,P30776); coded for by C. elegans CDNA
CEES77F"
/codon_start=1
/protein_id="AA79748.1"
/db_xref="PID:g1022971"
/db_xref="GI:1022971"
/translation="MCHPEFTICTISDASHTSSSRTSPSTPSTSPSPICIA
AKCSRCGVAAFYADVFLSKGLPSKVLAHWKELSKAQIFETDVEAVNEIMOP
SOKLAITGTHLLGICRYSRKTKYLLADCHAEFLKILVRSGLADQNPVLPFES
IQIYTGDNVLPEDEENHPACIQSRKLDITUKPQFTNYGIEPLEDDFG
ETAAVGPDYVLYLMEDEVNKMDELADKDAATTSNLFGREREPTFELNDAHAGNA
IFGDEAGGMDDDHVGSGSGFTDNNAMHMDYDESGNTMGETPLRSDPTNT
FRAPSPATSVATSVAMGHDTTSEYERQAQRAMQEKMRKRVRDVKMTGT
EEMGNMADFSLTLADLPPSKLMGKRSKSHAYMHHPGMIQFSKNKQIFRKYQ
SYLVYTHDDVDEKSEIKNALGLKEEGEELQQOQMDHVDQSSVYDDEVP
LDFDQLMDMQSMDINDVDDVQRPRLSPFAPWVEDEENSPKKRKRDEEKE
NOCIDIEKSPKTHALLQNTKLENGQVELDEMLKKGTSRKVAAAFYSLLCLKK
PIQDTEQKPEYGDIMIKAGPININTI"
31053. .34048
/gene="K08A8.1"
join(31053. .31236,31286. .31518,31569. .31712,31768. .32079,
32206. .32366,33479. .33583,33626. .33830,33875. .34048)
/gene="K08A8.1"
/note="similar to map kinase kinases; coded for by C.
elegans cDNA cm14d12; coded for by C. elegans cDNA cm04a9"
/codon_start=1
/protein_id="AA79746.1"
/db_xref="PID:g1022969"
/db_xref="GI:1022969"
/translation="MERDFDLGMRGGLGGEPIQMOPAPAPHPHSSRSDNHV
KNLMQAEENSYLTQGNRKNADLKELFVEDIGHSGCTVTKCRYSKIVMAVKTMP
RTNSYSEMRILMDLVQICRFDPCYIVRCFYGFTNFVRVCMCACTCDLRLIRI
KQPIPERIIGLSVSIKALHYLKTQHMHRDVCPSNLLDWSGVKLCDFGIAGRL
IESRASHKQACGLYMGKPELRDNDSDYIDISDWSFGVTLVELATGOYPYAGTFD
MNSKILNDEPPLDPKAFSPDFCQVESCLODRTPRPNYDMLLOHPFVYHKEIFD
VEWFDVWGDDSYAHNSNEISNLTOYAGSQQQIQLIQLENAFLKEQVKRENEQ
ERLIRNEMEQYADRFQEQAKREFDLOHRRESVSQINRSRQSTQTLARGNEVS
INGETWCPQHPHPPIAISDRSQIRGRDVRSLIIGFCSKTLKICKNRR"
1026 a 6242 c 6369 g 10710 t

11026

Best Local Similarity 94.4%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 TTAGGAATCTCCCACTT 20
||||| ||||| ||||| ||||| |||||

Db 29965 TTAGGAATCTCCCACTT 29982

RESULT 10
AC007360 80858 bp DNA PRI 05-JUN-1999
LOCUS Homo sapiens clone NH0104K07, complete sequence.
DEFINITION AC007360
ACCESSION 95001526
NID VERSION AC007360.2 GI:5001526
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 80858)
AUTHORS Waterston,R.H.
TITLE The sequence of Homo sapiens clone
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 80858)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (23-APR-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 3 (bases 1 to 80858)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
COMMENT On Jun 5, 1999 this sequence version replaced gi:4662676.
FEATURES
source
location 1..80858
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="NH0104K07"

BASE COUNT 25295 a 15886 c 16128 g 23549 t

ORIGIN

Query Match 82.0%; Score 16.4; DB 42; Length 80858;
Best Local Similarity 94.4%; Pred. No. 56;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 TTAGGAATCTCCCACT 19
||||| ||||| ||||| ||||| |||||

Db 3917 TTAGGAATCTCCCACT 3934

RESULT 11
LEU65668 6999 bp DNA PLN 01-JAN-1999
LOCUS Lycopersicon esculentum putative M1 copy 2 nematode-resistance gene.
DEFINITION LEU65668
ACCESSION U65668
NID 94090437
VERSION U65668.1 GI:4090437
KEYWORDS
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; potatoe; Lycopersicon.
REFERENCE 1 (bases 1 to 6999)
AUTHORS Williamson,V.M., Bodeau,J.P., Kaloshian,I., Yaghoobi,J. and Milligan,S.

TITLE Direct Submission
JOURNAL Submitted (31-JUL-1996) CEPRAP/Nematology, Univ.California, 1930 5th St., Davis, CA 95616, USA
FEATURES
source
location 1..6999
/organism="Lycopersicon esculentum"
/strain="Mottele"
/db_xref="taxon:4081"
/chromosome="6"
/map="Between GP79 and APS1"
misc_feature
2865..6483
/note="putative M1 gene; encodes putative nematode-resistance gene; Allele: R; exon/intron boundaries are not known"

BASE COUNT 2228 a 1070 c 1262 g 2439 t

ORIGIN

Query Match 80.0%; Score 16; DB 8; Length 6999;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTAGGAATCTCCCA 17
||||| ||||| ||||| ||||| |||||

Db 649 TTAGGAATCTCCCA 664

RESULT 12
LEU81378 51953 bp DNA PLN 19-AUG-1998
LOCUS Lycopersicon esculentum disease resistance gene homolog Mi-copy2
DEFINITION gene, complete cds; resistance gene pseudogene, complete sequence; disease resistance gene homolog Mi-copy1 gene, complete cds; and unknown gene.
ACCESSION U81378
NID 93426259
VERSION U81378.1 GI:3426259
KEYWORDS
SOURCE tomato.
ORGANISM Lycopersicon esculentum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; eudicotyledons; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids I; Solanales; Solanaceae; Solanum; potatoe; Lycopersicon.
REFERENCE 1 (bases 1 to 51953)
AUTHORS Rossi,M., Goggin,F.L., Milligan,S.B., Kaloshian,I., Ullman,D.E. and Williamson,V.M.
TITLE The nematode resistance gene M1 of tomato confers resistance against the potato aphid
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 95 (17), 9750-9754 (1998)
MEDLINE 98374252
REFERENCE 2 (bases 1 to 51953)
AUTHORS Bodeau,J.P., Kaloshian,I., Milligan,S. and Williamson,V.M.
TITLE Isolation and large-scale sequencing of bacterial artificial chromosomes (BACs) spanning the tomato M1 nematode-resistance locus
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 51953)
AUTHORS Bodeau,J.P., Kaloshian,I., Milligan,S. and Williamson,V.M.
TITLE Direct Submission
JOURNAL Submitted (07-DEC-1996) Nematology, CEPRAP, 1930 5th St., Davis, CA 95616, USA
FEATURES
source
location 1..51953
/organism="Lycopersicon esculentum"
/cultivar="Mottele"
/db_xref="taxon:4081"
/chromosome="6"
/map="between APS and GP79"
misc_feature
complement(1..1055)
/note="similar to Smcy/Retinoblastoma binding protein gene"
CDS
16312..19932

```

/codon_start=1
/product="disease resistance gene homolog Mi-copy2"
/protein_id="AAC32252.1"
/db_xref="PID:g3426260"
/db_xref="GI:3426260"
/translation="MAFICTYVQLSDFEQEDIMTRNRQEVNLLQSLDDVLTSLTSMDDCISLYHRSYKSDATMBDEOLDLNLNLYHLSKHAERKIFPGVTOYEVLOWNCGNIRDFGLNLGNCIKHEMVENVLPFLQMAERGVHFMEDQDDEDSRLSELDDEHNDORSLQLTHLLKIVPTELEVWHICYNLAKASTAEVGRFICKLLETSPDLREYITQOEHMTVPSTIPGARNHVMMEFLILLSDMPDFIHDHKLFDLLAHVGLTRFYVTLRDELEKLNKGNQNTNATLDLLELLKDLKHLKHYLKAPNSQCCFPMSDGPLFHLHMLHNDLDSNAYSISLKEEELVSOELEFIRSFAGAAEQGLKDIWAVLDVAYEAKDVDSIIVRDNGLHLIFSLPITIKKIKLKEISALDENIPKDRGLIVNSPKPVERKSLTDTKIIIVGFEETNLILKLTGSPADLDVISTGMPGSKTTLAKYVNDKSVSRHFDLRACVTDQGYDDKLLDITFQVSGSDSNLENIDVADKLRQLFGRKYLIVLDDVWMTTLDLTLPFPFAKNGSRILITTRKEVALHVKLNTDPLDLRLRDESVLLEDDKTFNESCDELDDVKGKIAENCKPLVADLTAAGTAGREKRSVYLVQESGLSFLINSVEVMKVIYSLDHPHLKPCLLHFAFWPKDPTLTYYLTVYLGAEVKEVTEKGEIEVVKIYMDLISLSSLVICFNEIGDILNFQIHDVHDFCLIKARKENLDRITSSAPDLPQITIDYDEEEHFGFNFMFDSNKKRHSGLHLYSLRNGDQLDSVSDAFHLRLRLIRVLDLPSLIMVNDLLEICMLNHLRYLRIYOVKYLPFSNLWNLESFVSNKGSIVLLPRILDLVKRLVSGACSFDDMDADESILIAKDTLENLIGELLISYKSDTMNIFKRPFLQVLFQELKESDYISTEQHWPFLKDLCTELETLCVGFKSSNTNHCSSVATNRPWDFHFNKELLYLDFLTSLSLSTARLPNLENLSLYDTITQGEWNGEDTFENLFLNRLTLTSKWEVGEESFPNLEKLLQECGKLEIPSPGDIYSLKIVKSPQLEDSALKIKKVAEDMRGNGEQLILQXKNIPLEF"
/27938..30928
/gene="resistance gene"
/pseudo
/27938..30928
/gene="resistance gene"
/codon_start=1
/pseudo
<41016..>42849
/gene="Lyt2"
41016..41432
/gene="Lyt2"
/note="putative transposase exon"
/number=1
/evidence=not_experimental
41851..42849
/gene="Lyt2"
/note="putative transposase exon"
/number=2
/evidence=not_experimental
45331..48945
/codon_start=1
/product="disease resistance gene homolog Mi-copy1"
/protein_id="AAC32253.1"
/db_xref="PID:g3426261"
/db_xref="GI:3426261"
/translation="MAFICTYVQLSDFEQEDIMTRKQFVENLLQPLDDVFTSLTSMDDCISLYHRSYKSDAEMDEOLDLNLNLYHLSKHAERKIFPGVTOYEVLOWNCGNIRDFGLNLGNCIKHEMVENVLPFLQMAERGVHFMEDQDDEDSRLSELDDEHNDORSLFLHLLKIVPTELEVWHICYNLAKASTAEVGRFICKLLETSPDLREYITQOEHMTVPSTIPGARNHVMMEFLILLSDMPDFIHDHKLFDLLAHVGLTRFYVTLRDELEKLNKGNQNTNATLDLLELLKDLKHLKHYLKAPNSQCCFPMSDGPLFHLHMLHNDLDSNAYSISLKEEELVSOELEFIRSFAGAAEQGLKDIWAVLDVAYEAKDVDSIIVRDNGLHLIFSLPITIKKIKLKEISALDENIPKDRGLIVNSPKPVERKSLTDTKIIIVGFEETNLILKLTGSPADLDVISTGMPGSKTTLAKYVNDKSVSRHFDLRACVTDQGYDDKLLDITFQVSGSDSNLENIDVADKLRQLFGRKYLIVLDDVWMTTLDLTLPFPFAKNGSRILITTRKEVALHVKLNTDPLDLRLRDESVLLEDDKTFNESCDELDDVKGKIAENCKPLVADLTAAGTAGREKRSVYLVQESGLSFLINSVEVMKVIYSLDHPHLKPCLLHFAFWPKDPTLTYYLTVYLGAEVKEVTEKGEIEVVKIYMDLISLSSLVICFNEIGDILNFQIHDVHDFCLIKARKENLDRITSSAPDLPQITIDYDEEEHFGFNFMFDSNKKRHSGLHLYSLRNGDQLDSVSDAFHLRLRLIRVLDLPSLIMVNDLLEICMLNHLRYLRIYOVKYLPFSNLWNLESFVSNKGSIVLLPRILDLVKRLVSGACSFDDMDADESILIAKDTLENLIGELLISYKSDTMNIFKRPFLQVLFQELKESDYISTEQHWPFLKDLCTELETLCVGFKSSNTNHCSSVATNRPWDFHFNKELLYLDFLTSLSLSTARLPNLENLSLYDTITQGEWNGEDTFENLFLNRLTLTSKWEVGEESFPNLEKLLQECGKLEIPSPGDIYSLKIVKSPQLEDSALKIKKVAEDMRGNGEQLILQXKNIPLEF"
/27938..30928
/gene="resistance gene"
/pseudo
/27938..30928
/gene="resistance gene"
/codon_start=1
/pseudo
<41016..>42849
/gene="Lyt2"
41016..41432
/gene="Lyt2"
/note="putative transposase exon"
/number=1
/evidence=not_experimental
41851..42849
/gene="Lyt2"
/note="putative transposase exon"
/number=2
/evidence=not_experimental
45331..48945
/codon_start=1
/product="disease resistance gene homolog Mi-copy1"
/protein_id="AAC32253.1"
/db_xref="PID:g3426261"
/db_xref="GI:3426261"
/translation="MAFICTYVQLSDFEQEDIMTRKQFVENLLQPLDDVFTSLTSMDDCISLYHRSYKSDAEMDEOLDLNLNLYHLSKHAERKIFPGVTOYEVLOWNCGNIRDFGLNLGNCIKHEMVENVLPFLQMAERGVHFMEDQDDEDSRLSELDDEHNDORSLFLHLLKIVPTELEVWHICYNLAKASTAEVGRFICKLLETSPDLREYITQOEHMTVPSTIPGARNHVMMEFLILLSDMPDFIHDHKLFDLLAHVGLTRFYVTLRDELEKLNKGNQNTNATLDLLELLKDLKHLKHYLKAPNSQCCFPMSDGPLFHLHMLHNDLDSNAYSISLKEEELVSOELEFIRSFAGAAEQGLKDIWAVLDVAYEAKDVDSIIVRDNGLHLIFSLPITIKKIKLKEISALDENIPKDRGLIVNSPKPVERKSLTDTKIIIVGFEETNLILKLTGSPADLDVISTGMPGSKTTLAKYVNDKSVSRHFDLRACVTDQGYDDKLLDITFQVSGSDSNLENIDVADKLRQLFGRKYLIVLDDVWMTTLDLTLPFPFAKNGSRILITTRKEVALHVKLNTDPLDLRLRDESVLLEDDKTFNESCDELDDVKGKIAENCKPLVADLTAAGTAGREKRSVYLVQESGLSFLINSVEVMKVIYSLDHPHLKPCLLHFAFWPKDPTLTYYLTVYLGAEVKEVTEKGEIEVVKIYMDLISLSSLVICFNEIGDILNFQIHDVHDFCLIKARKENLDRITSSAPDLPQITIDYDEEEHFGFNFMFDSNKKRHSGLHLYSLRNGDQLDSVSDAFHLRLRLIRVLDLPSLIMVNDLLEICMLNHLRYLRIYOVKYLPFSNLWNLESFVSNKGSIVLLPRILDLVKRLVSGACSFDDMDADESILIAKDTLENLIGELLISYKSDTMNIFKRPFLQVLFQELKESDYISTEQHWPFLKDLCTELETLCVGFKSSNTNHCSSVATNRPWDFHFNKELLYLDFLTSLSLSTARLPNLENLSLYDTITQGEWNGEDTFENLFLNRLTLTSKWEVGEESFPNLEKLLQECGKLEIPSPGDIYSLKIVKSPQLEDSALKIKKVAEDMRGNGEQLILQXKNIPLEF"

```

```

*
complement(49849..50313)
/note="similar to Arabidopsis EST PRL2-89B9T7"
/codon_start=1
/product="unknown"
/protein_id="AAC32254.1"
/db_xref="PID:g3426262"
/db_xref="GI:3426262"
/translation="MGCPKLOSLYFCROMTNDALYSIARNRNMIRFCILIEPOTPDYLLPBLDGFAGIYVQHCKEVORLSLSGLLTDVREYIGVHAKKLEMSLAFAGDSDIGLYLVSGESRUKLETRDCPDGDAALAAKLETRMSLMSNCSVSFEA"
BASE COUNT 17728 a 8428 c 8669 g 17115 t 13 others
ORIGIN

Query Match      80.08; Score 16; DB 8; Length 51953;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2 TTTAGGAATCTTCCCA 17
|||||
Db 14096 TTTAGGAATCTTCCCA 14111

RESULT 13
GGMITG/c 16412 bp DNA PRI 07-OCT-1996
LOCUS G.gorilla mitochondrial DNA. complete genome.
DEFINITION X93347
ACCESSION X93347
NID G1304307
VERSION X93347.1 GI:1304307
KEYWORDS 12S ribosomal RNA; 16S rRNA gene; 16S ribosomal RNA; 16S rRNA gene; COI gene; COII gene; COIII gene; control region; cytb gene; NADH1 gene; NADH2 gene; NADH3 gene; NADH4 gene; NADH5 gene; NADH6 gene; origin of replication; transfer RNA-Ala; transfer RNA-Arg; transfer RNA-Asn; transfer RNA-Asp; transfer RNA-Cys; transfer RNA-Gln; transfer RNA-Glu; transfer RNA-Gly; transfer RNA-His; transfer RNA-Ile; transfer RNA-Leu; transfer RNA-Lys; transfer RNA-Met; transfer RNA-Phe; transfer RNA-Pro; transfer RNA-Ser; transfer RNA-Thr; transfer RNA-Trp; transfer RNA-Tyr; transfer RNA-Val.

SOURCE
ORGANISM Mitochondrion Gorilla gorilla
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Gorilla.
REFERENCE 1 (bases 1 to 16412)
AUTHORS Xu,X. and Arnason,U.
TITLE A complete sequence of the mitochondrial genome of the Western lowland gorilla
JOURNAL Mol. Biol. Evol. 3, 691-698 (1996)
REFERENCE 2 (bases 1 to 16412)
AUTHORS Arnason,U.
TITLE Direct Submission
JOURNAL Submitted (16-NOV-1995) U. Arnason, Dept of Genetics, Division Evolutionary Molec. Systematics, University of Lund, Solvegatan 29, S-223 62 LUND, SWEDEN
FEATURES
Source
1. 16412
/organism="Gorilla gorilla"
/mitochondrion
/sub_species="G.g.gorilla"
/db_xref="taxon:9593"
/note="Western lowland gorilla"
1. 71
/product="transfer RNA-Phe"
72..1021
/gene="12S rRNA"
/product="12S ribosomal RNA"
72..1021
/gene="12S rRNA"
1022..1090
/product="transfer RNA-Val"
1091..12648

```

```
/gene="16S rRNA"
/product="16S ribosomal RNA"
1091..2648
/gene="16S rRNA"
2649..2723
/product="transfer RNA-Leu"
2726..3682
/gene="NADH1"
2726..3682
/gene="NADH1"
3682..3750
/product="transfer RNA-Ile"
complement(3748..3819)
/product="transfer RNA-Gln"
3821..3888
/product="transfer RNA-Met"
3889..4932
/gene="NADH2"
3889..4932
/gene="NADH2"
4931..4998
/product="transfer RNA-Tip"
complement(5006..5074)
/product="transfer RNA-Ala"
complement(5076..5148)
/product="transfer RNA-Asn"
5149..5182
/note="L strand"
complement(5180..5245)
/product="transfer RNA-Cys"
complement(5245..5310)
/product="transfer RNA-Tyr"
5326..6867
/gene="COI"
5326..6867
/gene="COI"
complement(6868..6936)
/product="transfer RNA-Ser"
6941..7008
/product="transfer RNA-Asp"
7009..7692
/gene="COII"
7009..7692
/gene="COII"
7719..7788
/product="transfer RNA-Lys"
7790..7996
/gene="ATPase8"
7790..7996
/gene="ATPase8"
7951..8631
/gene="ATPase6"
7951..8631
/gene="ATPase6"
8631..9413
/gene="COIII"
8631..9413
/gene="COIII"
9415..9482
/product="transfer RNA-Gly"
9483..9827
/gene="NADH3"
9483..9827
/gene="NADH3"
9829..9893
/product="transfer RNA-Arg"
9894..10190
/gene="NADH4L"
9894..10190
/codon_start=1
/gene="NADH4L"
/transl_table=2
/protein_id="CAA63723.1"

/db_xref="PID:e213277"
/db_xref="PID:g1304308"
/db_xref="GI:1304308"
/translation="MPLIYNIMLAFTISLLGMLVYRSH..VSSLLCLEGMMLSLFIMA
TMTLNTHTSLANIVPTIMLVFAACEAAVGLALLVSISNTYGLDYQNLLQC"
10184..11560
/gene="NADH4"
10184..11560
/gene="NADH4"
11562..11630
/product="transfer RNA-His"
11631..11689
/product="transfer RNA-Ser"
11690..11760
/product="transfer RNA-Leu"
11761..13572
/gene="NADH5"
11761..13572
/gene="NADH5"
complement(13573..14097)
/gene="NADH6"
complement(13573..14097)
/gene="NADH6"
complement(14098..14166)
/product="transfer RNA-Glu"
14171..15310
/gene="cyt b"
14171..15310
/gene="cyt b"
15312..15377
/product="transfer RNA-Thr"
complement(15381..15448)
/product="transfer RNA-Pro"
15449..16412
/note="control region"

BASE COUNT 5079 a 5039 c 2149 g 4145 t
ORIGIN

Query Match 80.0%; Score 16; DB 9; Length 16412;
Best Local Similarity 100.0%; Pred. No. 1.le-02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 TAGGAATCTTCCCACT 19
|||||
DB 1400 TAGGAATCTTCCCACT 1385

RESULT 14
GORMTC/c
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS

GORMTC 16364 bp DNA circular PRI 08-FEB-1999
Gorilla mitochondrial DNA, complete sequence.
D38114
9643689
D38114.1 GI:643689
cytochrome c oxidase subunit 1 (COI); cytochrome c oxidase subunit
3 (COIII); cytochrome c oxidase subunit 2 (COII); 12S rRNA; 16S
rRNA; ATPase subunit 6; ATPase subunit 8; NADH dehydrogenase
subunit 1; NADH dehydrogenase subunit 2; NADH dehydrogenase subunit
3; NADH dehydrogenase subunit 4; NADH dehydrogenase subunit 5;
cytochrome b; cytochrome c oxidase subunit I; cytochrome c oxidase
subunit II; cytochrome c oxidase subunit III; trRNA-Ala; trRNA-Arg;
trRNA-Asn; trRNA-Asp; trRNA-Cys; trRNA-Gln; trRNA-Gly; trRNA-His;
trRNA-Ile; trRNA-Leu; trRNA-Leu(CUN); trRNA-Lys; trRNA-Phe;
trRNA-Pro; trRNA-Ser(AGY); trRNA-Ser(UCN); trRNA-Thr; trRNA-Trp;
trRNA-Tyr; trRNA-Val.
Gorilla gorilla mitochondrion DNA.
Mitochondrion Gorilla gorilla
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates. Catarrhini; Homnidae; Gorilla.
1 (base. 1 to 16364)
```


AUTHORS Hayasaka, K.
 TITLE Direct Submission
 JOURNAL Submitted (02-SEP-1994) to the DDBJ/EMBL/GenBank databases. Kenji Hayasaka, National Institute of Genetics, Human Genetics; 1,111 Yata, Mishima, Shizuoka 411, Japan (E-mail:khayasak@dbj.nig.ac.jp, Tel:81-559-75-0771(ex.568))
 REFERENCE 2 (sites)
 AUTHORS Hixson, J.E. and Brown, W.M.
 TITLE A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications
 JOURNAL Mol. Biol. Evol. 3 (1), 1-18 (1986)
 MEDLINE 88174329
 REFERENCE 3 (sites)
 AUTHORS Foran, D.R., Hixson, J.E. and Brown, W.M.
 TITLE Comparisons of ape and human sequences that regulate mitochondrial DNA transcription and D-loop DNA synthesis
 JOURNAL Nucleic Acids Res. 16 (13), 5841-5861 (1988)
 MEDLINE 88289338
 REFERENCE 4 (sites)
 AUTHORS Horai, S., Satta, Y., Hayasaka, K., Kondo, R., Inoue, T., Ishida, T., Hayashi, S. and Takahata, N.
 TITLE Man's place in Hominoidea revealed by mitochondrial DNA genealogy
 JOURNAL J. Mol. Evol. 35 (1), 32-43 (1992)
 MEDLINE 92389366
 REMARK Erratum: [[published erratum appears in J Mol Evol 1993 Jul;37(1):89]]
 REFERENCE 5 (sites)
 AUTHORS Horai, S., Hayasaka, K., Kondo, R., Tsugane, K. and Takahata, N.
 TITLE Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 92 (2), 532-536 (1995)
 MEDLINE 95132634
 REFERENCE 6 (bases 1 to 16364)
 AUTHORS Horai, S., Kazuo, T., Hayasaka, K., Naoyuki, T. and Rumi, K.
 JOURNAL Unpublished (1995)
 FEATURES
 source
 1..16364
 /organism="Gorilla gorilla"
 /mitochondrion
 /db_xref="taxon:9593"
 tRNA 1..71
 /product="tRNA-Phe"
 rRNA 72..1020
 /product="12S rRNA"
 tRNA 1021..1089
 /product="tRNA-Val"
 rRNA 1090..2647
 /product="16S rRNA"
 tRNA 2648..2722
 /product="tRNA-Leu(UUR)"
 mat_peptide 2725..3678
 /product="NADH dehydrogenase subunit 1 (ND1)"
 tRNA 3681..3749
 /product="tRNA-Ile"
 tRNA complement(3747..3818)
 /product="tRNA-Gln"
 tRNA 3820..3887
 /product="tRNA-Met"
 mat_peptide 3888..4928
 /product="NADH dehydrogenase subunit 2 (ND2)"
 tRNA 4930..4997
 /product="tRNA-Trp"
 tRNA complement(5005..5073)
 /product="tRNA-Ala"
 tRNA complement(5075..5147)
 /product="tRNA-Asn"
 tRNA complement(5179..5244)
 /product="tRNA-Cys"
 tRNA complement(5244..5309)
 /product="tRNA-Tyr"
 mat_peptide 5325..6863
 /product="cytochrome c oxidase subunit 1 (COI)"
 complement(6866..6937)
 /product="tRNA-Ser(UCN)"
 tRNA 6939..7006
 /product="tRNA-Asp"
 CDS 7007..7690
 /codon_start=1
 /transl_table=2
 /product="cytochrome c oxidase subunit 2 (COII)"
 /protein_id="BAA07303.1"
 /db_xref="PID:d1007881"
 /db_xref="PID:g704436"
 /db_xref="GI:704436"
 /translation="MAHAAQVGLQDATSPIMEELIIFHDHALMIIFLIGFLVLYALFL TLTKLTNNNISDAQEMETWILPAIILVLIAPSLRILYIMTDEINDPSFTIKSIGH QWYIYETDYGGLIFNSYMLPPLFLEPGDLRLDDNRWLVPVEAPVRMITSDQVL HSAVPTLGLKTDALPGRNQTTTATPRGVYVYGCSEICGANHSMPVILELIPKLI FEMGPFVAL"
 tRNA 7717..7786
 /product="tRNA-Lys"
 CDS 7788..7994
 /codon_start=1
 /transl_table=2
 /product="ATPase subunit 8 (ATPase8)"
 /protein_id="BAA07304.1"
 /db_xref="PID:d1007882"
 /db_xref="PID:g704437"
 /db_xref="GI:704437"
 /translation="MPOLENTVWPTMIAPMLLTLELTOLKVLNTNYHLPPLPKTMKM KNFCKPWPWKTKIYSLHLPQS"
 mat_peptide 7949..8626
 /product="ATPase subunit 6 (ATPase6)"
 mat_peptide 8629..9411
 /product="cytochrome c oxidase subunit 3 (COIII)"
 tRNA 9413..9480
 /product="tRNA-Gly"
 mat_peptide 9481..9825
 /product="NADH dehydrogenase subunit 3 (ND3)"
 tRNA 9827..9891
 /product="tRNA-Arg"
 CDS 9892..10188
 /codon_start=1
 /transl_table=2
 /product="NADH dehydrogenase subunit 4L (ND4L)"
 /protein_id="BAA07305.1"
 /db_xref="PID:d1007883"
 /db_xref="PID:g704438"
 /db_xref="GI:704438"
 /translation="MPLIYNNIMLAFTISLLGMLVYRSHLSSLLCLGMMLSLFMA TLTNTHTFLLANIVPIITMLVFAACEAAVGLALLVSINTYGLDYVQNLNLIQC"
 mat_peptide 10182..11558
 /product="NADH dehydrogenase subunit 4 (ND4)"
 tRNA 11560..11628
 /product="tRNA-His"
 tRNA 11629..11687
 /product="tRNA-Ser(AGY)"
 tRNA 11688..11758
 /product="tRNA-Leu(CUN)"
 CDS 11759..13570
 /codon_start=1
 /transl_table=2
 /product="NADH dehydrogenase subunit 5 (ND5)"
 /protein_id="BAA07306.1"
 /db_xref="PID:d1007884"
 /db_xref="PID:g704439"
 /db_xref="GI:704439"
 /translation="MTMTATMTLTSLTSLIPILTTTINPNKSSPHYVKSIVASTF IISLFTPTMFLCQDQEAIISSWHWATTQTQLSLSKFLDYFSMMFIPALFVTVSIME FSLWMNSDPNIQOFFKYLILITLMLVITANNLFOLFLGWEGVMSFLIGWMA RTDANTAQAQVAILNRIIGDIGFILALAWFLHSHNSWDPQOSSLTNPNLPLGLFL AAAGKSAQLGLHPWLPSPMGPTVPSALLHSSYVYAGVLLIRFHLAENNSLAQTL TLCIGAITTLFAAVCALQTQNDIKIVAFSTSSQLGMLVATIGIGQHLAFLHCTHAF FRAMLEWCSSIIHNNEQDIRKMGGLKAMPLTSTSLAIGSLAMGMPFLTGFSYK DLIETANNSHTNAWLSIILATSLTSAYSTRMILTATGQPRFPTFANINENYSTL

```

LNPKRITGSLRAGPFTNNILPTSPQMTIDLYLKLTPALSTILGLLTALDNLVLT
NKLMKPHPTFFSNMGLFYPNTHRTIPYLGLLMSQNLPLLLLDLIWLEKLPKTI
SQHQISAITTSQKGLIKLYFLSFEPFLILLIILLIT"
CDS      complement(13571..14095)
        /note="unnamed protein product"
        /codon_start=1
        /transl_table=2
        /protein_id="BAA07307.1"
        /db_xref="PID:d1007885"
        /db_xref="PID:g1213359"
        /db_xref="GI:1213359"
        /translation="MTVLFLLSLGLVMGFVGFSSKPSPIYGLVLIVSGWGCAIIL
NCGGYNGLWFLYLGGRZVFCYTTAMAIGEPYPAWGSQVEVLVSLVGLAMEVGL
VLWVKYDGVVVVNFNNYGSWMI:EGEGSLIREDFIGAGALIDYGRNLVVYTGWTL
FVGVIIVIEIARGN"
tRNA     complement(14096..14164)
        /product="tRNA-Gln"
        14169..15308
mat_peptide
        /product="cytochrome b"
        15310..15377
tRNA     /product="tRNA-Thr"
        complement(15379..15446)
        /product="tRNA-Pro"
        15447..16364
D-loop   5059 a 5022 c 2160 g 4123 t
ORIGIN

Query Match      80.0%; Score 16; DB 9; Length 16364;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      4 TAGGAATCTTCCCACT 19
        |||||
Db      1399 TAGGAATCTTCCCACT 1384

RESULT 15
HSAC000380 164296 bp DNA HTG 26-MAR-1997
LOCUS Homo sapiens chromosome 3 clone pDJ70111, WORKING DRAFT SEQUENCE, 2
DEFINITION
unordered pieces.
ACCESSION AC000380
NID g1907283
VERSION AC000380.1 GI:1907283
KEYWORDS HTG; HTGS-PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 164296)
AUTHORS Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J.,
Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.,
Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
Wilson,R. and Burbee,D.
TITLE Unpublished
JOURNAL 2 (bases 1 to 164296)
REFERENCE Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J.,
AUTHORS Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.,
Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
Wilson,R. and Burbee,D.
TITLE Direct Submission
JOURNAL Submitted (26-MAR-1997) Genome Science and Technology Center,
REFERENCE University of Texas Southwestern Medical Center at Dallas, 5323
AUTHORS Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
```

```

* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 7834: contig of 7834 bp in length
* 7835 164296: contig of 156462 bp in length.
FEATURES
        source
        1..164296
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="pDJ70111"
            /chromosome="3"
BASE COUNT 38527 a 43485 c 42989 g 39285 t 10 others
ORIGIN

Query Match      80.0%; Score 16; DB 34; Length 164296;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      5 AGGAATCTTCCCACTT 20
        |||||
Db      25193 AGGAATCTTCCCACTT 25208

Search completed: September 18, 1999, 15:48:58
Job time: 1418 sec
```

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 05:27:18 ; Search time 213.04 Seconds
(without alignments)
23.488 Million cell updates/sec

Title: US-09-037-472-4
Perfect score: 20
Sequence: 1 GTTTAGGAATCTTCCCACTT 20

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	20	100.0	20	T13886	Primer for detecti
2	20	100.0	20	V32392	Interleukin-1-beta
3	20	100.0	20	V60233	PCR primer used to
4	20	100.0	20	X16616	Interleukin 1 (441
5	20	100.0	14690	X22303	Human IL-1ra BAC c
6	20	100.0	10620	X02996	Human IL-1ra BAC c
7	16	80.0	9870	V16457	Wild tomato Mi res
8	16	80.0	9870	V13935	Tomato Mi resistan
9	16	80.0	51952	V26084	Tomato pest resist
10	15.8	79.0	3186	Q14631	Plasmid pTM72 inse
11	15.8	79.0	3890	Q71539	Brain low Km, CAMP
12	15.8	79.0	3186	T34375	Plasmid pTM72 (ATC
13	15.2	76.0	10798	Q28505	Tomato vacuolar in
14	15.2	76.0	2199	Q28504	Tomato vacuolar in
15	15.2	76.0	10965	Q28507	Tomato acid invert
16	15.2	76.0	2223	Q39051	Tomato acid invert
17	15.2	76.0	2339	Q63892	Tomato acid invert
18	15.2	76.0	2372	Q94259	Invertase gene of
19	15.2	76.0	2339	T01339	Tomato acid invert
20	15.2	76.0	9936	V34683	Arabidopsis thalia
21	15.2	76.0	18613	V74423	Staphylococcus aur
22	15.2	76.0	11764	X12985	Enterococcus faeca
23	15.2	76.0	403	X20936	Polynucleotide seq
24	15	75.0	110000	V21209_12	Continuation (13 o
25	14.8	74.0	8760	Q73473	Porcine pro-interl
26	14.8	74.0	2101	Q83528	P. falciparum EBL-
27	14.8	74.0	411	T23442	Human gene signatu
28	14.8	74.0	2101	T72896	Plasmodium ebl-2 g
29	14.8	74.0	7295	T72883	Plasmodium var-1 g
30	14.8	74.0	1064	T98622	Branched chain ami
31	14.8	74.0	8160	V99559	Arabidopsis lysine
32	14.8	74.0	251	X12416	Human biallelic po
33	14.4	72.0	2474	T70840	Mouse apoptosis in
34	14.4	72.0	2676	V55042	Murine HIP-1 codi
35	14.4	72.0	316	V90103	EST clone DGL7. Ne
36	14.2	71.0	9737	N60140	Sequence of ARV-2
37	14.2	71.0	1649	Q14638	Human temporal lob
38	14.2	71.0	1811	Q14635	Clone associated w
39	14.2	71.0	1672	Q14637	Clone associated w
40	14.2	71.0	11866	N40141	Sequence of human
41	14.2	71.0	11873	N40176	Sequence of human
42	14.2	71.0	9326	Q25051	DNA complementary
43	14.2	71.0	9737	Q31936	ARV-2 proviral DNA

44 14.2 71.0 27 1 Q65945 COL2A1 5'-primer (

c 45 14.2 71.0 2625 1 Q99851 Yeast TTH1 gene. I

ALIGNMENTS

RESULT 1

TI3886 standard; DNA; 20 BP.
ID T13886;
AC T13886;
DT 19-OCT-1997 (first entry)
DE Primer for detecting genetic predisposition to periodontal disease.
KW Periodontal disease; gingivitis; periodontitis; polymorphism;
KW Interleukin-1 alpha; IL-1A; interleukin-1 beta; IL-1B; primer; PCR;
KW Polymerase chain reaction; ss.
OS Synthetic.
PN WO9706180-A1.
PD 20-FEB-1997.
PF 02-AUG-1996; U12455.
PR 03-AUG-1995; US-510696.
PA (KORN/) KORNMAN K S.
PA (MEDI-) MEDICAL SCI SYSTEMS INC.
PI Duff GW, Kornman KS;
DR WPI: 97-154207/14.
PT Identification of patient's genetic polymorphism pattern - allows
PT Prediction of increased periodontal disease severity
PS Claim 3; Page 25; 35pp; English.
CC PCR primers (T70316 and T13884-88) can be used in a new method
CC for identifying a genetic predisposition to periodontal disease by
CC detecting the presence of DNA polymorphisms in the gene sequences
CC for interleukin-1 alpha (IL-1A) and interleukin-1 beta (IL-1B).
CC Alleles associated with severe disease were identified as IL-1A
CC allele 2 together with IL-1B (Taql) allele 2. The primers are used
CC to amplify DNA from a blood or tissue sample and products are
CC subjected to restriction digestion to determine the polymorphism
CC pattern of the patient. A single base variation (C/T) polymorphism
CC at IL-1B base -511 can be identified using primers (T13885 and
CC T13886) corresponding to bases -702 to -682 and -417 to -397,
CC respectively. The single base variation completes an Aval site
CC on allele 1 (C) and a Bsu361 site on allele 2 (T).
SQ Sequence 20 BP; 4 A; 5 C; 3 G; 8 T;

Query Match 100.0%; Score 20; DB 1; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.23; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTTAGGAATCTTCCCACTT 20

Db 1 GTTTAGGAATCTTCCCACTT 20

RESULT 2

V32392 standard; DNA; 20 BP.
ID V32392;
AC V32392;
DT 11-SEP-1998 (first entry)
DE Interleukin-1 beta primer 2 (-417/-398).
KW IL-1-beta; genetic polymorphism; PCR; primer; amplification;
KW sight threatening diabetic retinopathy; interleukin-1-alpha;
KW Interleukin-1-beta; interleukin-1RN; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9815653-A1.
PD 16-APR-1998.
PF 09-OCT-1997; G02790.
PR 10-OCT-1996; GB-021129.
PA (DUFF/) DUFF G.
PA (RENN/) RENNIE I.
PA (RICH/) RICHARDSON R.
PI Duff G, Rennie I, Richardson R;
DR WPI: 98-240835/21.

PT Predicting increased risk of sight-threatening diabetic retinopathy
 PT - comprises identifying genetic polymorphism pattern for genes
 PT IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
 PT symptoms occur
 PS Claim 2; Page 33; 41pp; English.
 CC Interleukin-1-beta (IL-1-beta) primers 2 and 1 (V32391) were used to
 CC amplify the IL-1-beta gene region to identify single base variation
 CC polymorphism of C/T at base 511. The invention claims to provide a
 CC method for predicting the risk of sight threatening diabetic retinopathy.
 CC The method involves isolating DNA from a patient and determining the DNA
 CC polymorphism pattern of the genes that code for interleukin-1-alpha,
 CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
 CC identified is then compared with controls of known DNA polymorphism
 CC patterns thereby identifying patients carrying a genetic polymorphism
 CC associated with increased risk of sight threatening diabetic retinopathy.
 CC The method may be able to identify diabetic patients at risk before the
 CC clinically detectable disorders occur. Polymorphism pattern
 CC determination of IL genes involved PCR reactions using primers V32389-
 CC V32398. The method is also claimed to be useful in conjunction with
 CC identification of other genes associated with sight threatening diabetic
 CC retinopathy in genomic DNA and therefore, in identifying diabetic
 CC patients expressing multiple risk patterns.
 CC Sequence 20 BP; 4 A; 5 C; 3 G; 8 T;
 SQ

Query Match 100.0%; Score 20; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTCCCACTT 20

Db 1 GTTTAGGAATCTCCCACTT 20

RESULT 3

V60233
 ID V60233 standard; DNA; 20 BP.
 AC V60233;
 DT 25-NOV-1998 (first entry)
 DE PCR primer used to amplify interleukin-1B (IL-1B).
 KW Interleukin-1B; IL-1B; predisposition; coronary artery disease;
 KW screen; PCR primer; ss.
 OS Synthetic.
 PN W09840517-A1.
 PD 17-SEP-1998.
 PF 09-MAR-1998; U04725.
 PR 10-MAR-1997; US-813456.
 PA (MEDI-) MEDICAL SCI SYSTEMS INC.
 PI Crossman DC, Duff GW, Francis SE;
 DR WPI: 98-520829/44.
 PT Detection of predisposition to coronary artery disease - by
 PT comparative measurement of levels of expression of alleles from the
 PT interleukin 1 locus
 PS Claim 6; Page 15; 22pp; English.
 CC PCR primers V60232-33 were used to amplify alleles associated with
 CC interleukin-1B (IL-1B). The specification describes a method for
 CC determination of a patient's predisposition to coronary artery
 CC disease, the method comprises comparing an allele with a second
 CC allele which is predictive of coronary artery disease, where
 CC similarity between the first and second alleles indicates a
 CC predisposition to coronary artery disease. The method is
 CC used to genotype an individual's interleukin (IL)-1 loci, the
 CC overexpression of which correlates with coronary artery disease. The
 CC method is used to screen a patients' predisposition to coronary
 CC artery disease.
 CC Sequence 20 BP; 4 A; 5 C; 3 G; 8 T;
 SQ

Query Match 100.0%; Score 20; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTCCCACTT 20

Db 1 GTTTAGGAATCTCCCACTT 20

RESULT 4

X16616
 ID X16616 standard; DNA; 20 BP.
 AC X16616;
 DT 29-APR-1999 (first entry)
 DE Interleukin 1 (44112332) haplotype PCR primer #10.
 KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;
 KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;
 KW Graves disease; systemic lupus erythematosus; lichen sclerosis;
 KW ulcerative colitis; PCR primer; ss.
 OS Synthetic.
 OS Homo sapiens.
 PN W09854359-A1.
 PD 03-DEC-1998.
 PF 21-MAY-1998; G01481.
 PR 29-MAY-1997; GB-011040.
 PA (CAMP/) CAMP N J.
 PA (COXA/) COX A.
 PA (DGIO/) DE GIOVINE F S.
 PA (DUFF/) DUFF G.
 PI Camp NJ, Cox A, De Giovine FS, Duff G;
 DR WPI: 99-080814/07.

PT New method of determining a patient's susceptibility to inflammatory
 PT disorders - by detecting the presence of an IL-1 (44112332)
 PT haplotype, useful in designing treatment strategies that modulate
 PT the activity of proteins produced by the IL-1 gene cluster
 PS Claim 3; Page 33; 49pp; English.
 CC A method has been developed for determining a patient's susceptibility
 CC to an inflammatory disorder. The method comprises the detection of an
 CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the
 CC patient, where its presence indicates susceptibility to an inflammatory
 CC disorder. X16607 to X16631 represent PCR primer used in the method for
 CC detecting the IL-1 (44112332) haplotype. The method provides kits for
 CC the early prediction of a patient's susceptibility to inflammatory
 CC disorders, including coronary artery disease, osteoporosis, nephropathy
 CC in diabetes mellitus, alopecia areata, Graves disease, systemic lupus
 CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of
 CC alleles of the haplotype can be applied to particular inflammatory
 CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,
 CC psoriasis, and insulin dependent diabetes. The identification of a
 CC disease-associated haplotype enables determination of which alleles are
 CC causative, and this information is useful in designing treatment
 CC strategies, including gene therapy and treatment using various agents
 CC that modulate the activity of proteins produced by the IL-1 gene cluster.
 CC Some alleles from the IL-1 gene cluster are associated with particular
 CC inflammatory diseases, and insufficient IL-1 production appears to act
 CC centrally in the pathology of these diseases. Therefore, the use of IL-1
 CC gene clusters is useful in determining genetic susceptibility to
 CC inflammatory diseases, including those with a multifactorial etiology
 CC with a polygenic component.
 CC Sequence 20 BP; 4 A; 5 C; 3 G; 8 T;
 SQ

Query Match 100.0%; Score 20; DB 1; Length 20;
 Best Local Similarity 100.0%; Pred. No. 0.23;
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTCCCACTT 20

Db 1 GTTTAGGAATCTCCCACTT 20

RESULT 5

X22303
 ID X22303 standard; DNA; 14690 BP.
 AC X22303;
 DT 22-JUN-1999 (first entry)
 DE Human IL-1ra BAC contiguous DNA sequence 96.
 KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;

KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
PN WO9906426-A1.
PD 11-FEB-1999.
PF 03-AUG-1998; U16102.
PR 02-JUL-1998; US-091650.
PR 04-AUG-1997; US-054646.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Pan Y.
DR WPI; 99-153692/13.
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
PS Example 5; Figure 4; 226pp; English.
CC X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC containing alternatively spliced forms of human IL-1ra. Such fragments
CC are used in the method of the invention which describes the isolation of
CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC member of the cytokine superfamily that is expected to inhibit
CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC also bind to a new receptor so could regulate other cellular processes
CC associated with acute or chronic inflammation, e.g. asthma, chronic
CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent
CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 14690 BP; 3810 A; 3245 C; 3299 G; 4327 T;

Query Match 100.0%; Score 20; DB 1; Length 14690;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAGGAATCTCCCACTT 20
|||||
Db 10216 GTTAGGAATCTCCCACTT 10235

RESULT 6
X02996
ID X02996 standard; DNA; 10620 BP.
AC X02996;
DT 22-JUN-1999 (first entry)
DE Human IL-1ra BAC contiguous DNA sequence 41.
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
PN WO9906426-A1.
PD 11-FEB-1999.
PF 03-AUG-1998; U16102.
PR 02-JUL-1998; US-091650.
PR 04-AUG-1997; US-054646.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Pan Y.
DR WPI; 99-153692/13.
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
PS Example 5; Figure 3; 226pp; English.
CC X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC containing alternatively spliced forms of human IL-1ra. Such fragments
CC are used in the method of the invention which describes the isolation of
CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC member of the cytokine superfamily that is expected to inhibit
CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC also bind to a new receptor so could regulate other cellular processes
CC associated with acute or chronic inflammation, e.g. asthma, chronic
CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent

CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 10620 BP; 2746 A; 2407 C; 2411 G; 3049 T;

Query Match 100.0%; Score 20; DB 1; Length .0620;
Best Local Similarity 100.0%; Pred. No. 0.32;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTAGGAATCTCCCACTT 20
|||||
Db 6056 GTTAGGAATCTCCCACTT 6075

RESULT 7
V16457
ID V16457 standard; DNA; 9870 BP.
AC V16457;
DT 07-JUL-1998 (first entry)
DE Wild tomato Mi resistance gene.
KW AFLP marker PM14; root knot nematode; Mi; resistance gene;
KW wild tomato; pathogen resistance; pest resistance; aphid;
KW Meloidogyne incognita; ss.
OS Lycopersicon peruvianum.
FH Key Location/Qualifiers
FT CDS 3263..7111
FT /*tag= a
FT /*note= "open reading frame 1"
FT 3491..7111
FT /*tag= b
FT /*note= "open reading frame 2"
FT 1936..3241
FT /*tag= c
FT /*number= 1
FT 3305..3379
FT /*tag= d
FT /*number= 2
FT misc_feature 6921..7034
FT /*tag= e
FT /*note= "AFLP marker PM14"

WO9806750-A2.
19-FEB-1998.
PD 08-AUG-1997; E04340.
PF 16-MAY-1997; EP-401101.
PR 09-AUG-1996; EP-401764.
PA (KEYG-) KEYGENE NV.
PI Simons G, Vos P, Wijbrandi J, Zabeau M;
DR WPI; 98-159460/14.
DR P-PSDB; W53582. W53583.
PT New isolated Meloidogyne incognita resistance gene - used to
PT transform plants to provide resistance to pathogens or pests,
PT particularly nematodes or aphids.
PS Claim 1; Fig 5; 79pp; English.
CC The present sequence is an approximately 9.9 kb sequence located
CC around the wild tomato, Lycopersicon peruvianum (PI 128657), AFLP
CC marker PM14, which comprises the root knot nematode Meloidogyne
CC incognita (Mi) resistance gene. The gene can be used to transform
CC plants to provide resistance to plant pathogens or pests,
CC particularly nematodes, e.g. root-knot nematodes, or aphids.
SQ Sequence 9870 BP; 3249 A; 1563 C; 1686 G; 3372 T;

Query Match 80.0%; Score 16; DB 1; Length 9870;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTAGGAATCTCCCA 17
|||||
Db 1275 TTTAGGAATCTCCCA 1290

RESULT 8
V13935

```

ID V13935 standard; DNA; 9870 BP.
AC V13935;
DT 20-JUL-1998 (first entry)
DE Tomato Mi resistance gene.
KW Mi resistance gene; tomato; nematode resistance;
KW Meloidogyne incognita; crop protection; transgenic plant; ds.
OS Lycopersicon esculentum.
FH Key Location/Qualifiers
FT CDS 3491..7111
FT FT /*tag= a
FT FT /*note= "Claim 4"
PN EP-823481-A1.
PD 11-FEB-1998.
PF 09-AUG-1996; 401764.
PR (KEYG-) KEYGENE NV.
PA Simons G, Vos P, Wijnbrandi J, Zabeau M;
PI WPI: 98-112270/11.
DR P-PSDB: W47080.
PT Tomato Mi resistance gene - for producing nematode-resistant
transgenic plants
PS Claim 1; Fig 5A-D; 48pp; English.
CC This nucleotide sequence comprises the tomato Mi resistance gene
CC the confers resistance against nematodes. It codes for a 1207
CC amino acid protein (see W27080). The Mi resistance gene was cloned
CC from a tomato genotype that is resistant to Meloidogyne incognita
CC using a positional cloning strategy comprising: identification of
CC molecular markers linked to the Mi resistance gene; construction of
CC a high mol.wt. yeast artificial chromosome (YAC) library; physical
CC mapping of the molecular markers on the YAC clones and YAC contig
CC building; construction of a cosmid library of the YAC clones
CC harbouring the linked molecular markers; physical fine mapping and
CC cosmid contig building; genetic characterisation of tomato mutants
CC susceptible to root-knot nematodes; transformation of susceptible
CC plants with the cosmid forming the contig; and complementation
CC analysis. The invention further relates to genetically transformed
CC plants which are resistant to nematodes (especially Meloidogyne
CC incognita), and to probes and primers useful for identification of
CC resistance genes.
SQ Sequence 9870 BP; 3249 A; 1563 C; 1686 G; 3372 T;

Query Match 80.0%; Score 16; DB 1; Length 9870;
Best Local Similarity 100.0%; Pred. No. 29;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTAGGAATCTTCCCA 17
| | | | | | | | | | | | | | | | | | | | |
DB 1275 TTTAGGAATCTTCCCA 1290

RESULT 9
ID V26084 standard; cDNA; 51952 BP.
AC V26084;
DT 07-JUN-1999 (first entry)
DE Tomato pest resistance Mi gene (copy 1).
KW pest resistance; nematode resistance; disease resistance; Mi gene;
KW tomato; transgenic plant; crop protection; biological control; ss.
OS Lycopersicon esculentum.
FH Key Location/Qualifiers
FT promoter 10071..14691
FT FT /*tag= a
FT FT /*note= "copy 2 promoter; specifically claimed in
FT FT nucleic acid construct of Claim 9"
FT CDS 16084..19932
FT FT /*tag= b
FT FT /*note= "Mi gene (copy 2); encodes W55975"
FT FT 42798..44461
FT FT /*tag= c
FT FT /*note= "copy 1 gene promoter"
FT CDS 45102..48944
FT FT /*tag= d

```

```

FT W09815171-A1.
PN 16-APR-1998.
PD 09-OCT-1997; U18802.
PF 10-OCT-1996; US-028191.
PR (REGC ) UNIV CALIFORNIA.
PA Bodeau J, Kaloshian I, Milligan S, Williamson VM,
PI Yaghoobi J;
DR WPI: 98-240529/21.
DR P-PSDB: W55974, W55975.
PT Nucleic acids encoding Mi polypeptide(s) conferring nematode
PT resistance - useful to produce transgenic plants resistant to these
PT and other pests, and in marker-aided selection to assess cultivars
PT for resistance
PS Claim 11; Page 24-39; 55pp; English.
CC This is the nucleotide sequence of the tomato Mi locus associated
CC with nematode resistance, obtained from a bacterial artificial
CC chromosome (BAC). Mi was localised by genetic analysis to a region
CC of the tomato genome of about 65 kb. DNA corresponding to this
CC region was cloned into BAC vectors. Sequence analysis of a 52 kb
CC BAC3 insert identified 3 open reading frames, one of which is
CC probably a pseudogene. By RNA blot analysis, transcripts of
CC approximately 4 kb corresponding to copy 1 and copy 2 were found in
CC both resistant and susceptible tomato roots and in leaves of
CC resistant but not susceptible plants. cDNA sequences corresponding
CC to full-length transcripts of copy 1 (see. V26082) and copy 2 (see
CC V26083) were obtained. The encoded polypeptides (see W55974-75)
CC are 91% identical and contain structural features similar to known
CC plant resistance genes (R genes) of the nucleotide binding site/
CC leucine-rich repeat (NBS/LRR) family. A recombinant expression
CC cassette comprising an Mi polynucleotide and an operably linked
CC plant promoter can be used to enhance nematode resistance in plants
CC especially tomatoes (claimed). Transgenic plants can also be
CC constructed using a Mi promoter with heterologous genes; the Mi
CC promoters can be used to express a variety of genes in the same
CC temporal and spatial patterns and at similar levels to resistance
CC genes.
SQ Sequence 51952 BP; 17728 A; 8444 C; 8647 G; 17111 T;

Query Match 80.0%; Score 16; DB 1; Length 51952;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TTTAGGAATCTTCCCA 17
| | | | | | | | | | | | | | | | | | | | |
DB 14096 TTTAGGAATCTTCCCA 14111

RESULT 10
ID Q14631 standard; DNA; 3186 BP.
AC Q14631;
DT 30-JAN-1992 (first entry)
DE Plasmid pTM72 insert encoding a human cAMP phosphodiesterase.
KW Cyclic nucleotide; PDE; glioblastoma cell; rolipram-sensitive; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 139..2348
FT FT /*tag= a
PN W09116457-A.
PD 31-OCT-1991.
PF 19-APR-1991; U02714.
PR 20-APR-1990; US-511715.
PA (COLD-) COLD SPRING HARBOR.
PI Wigler MH, Colicelli JJ;
DR WPI: 91-339841/46.
DR P-PSDB: R14843.
PT Complementary screening for genes and prods. - e.g. RAS protein
PT and cAMP, that modify, complement or suppress genetic defect and
PT correct associated phenotypic alteration
PS Claim 15; Page 91; 169pp; English.
CC Plasmid pTM72 contains a human glioblastoma cDNA which encodes a

```

CC rolipram-sensitive cAMP PDE. The sequence is very closely related
 CC to, but distinct from, the rat pRATPD cDNA insert (see Q14624).
 CC N.B. This sequence is SEQ ID NO. 22 in the specification but is
 CC referred to as SEQ ID NO. 23 in the text, e.g. on page 32.
 SQ Sequence 3186 BP; 927 A; 743 C; 734 G; 782 T;

Query Match 79.0%; Score 15.8; DB 1; Length 3186;
 Best Local Similarity 89.5%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TTTAGGAATCTCCCACTT 20
 Db 2706 TTTATGAATCTTCTCACTT 2724

RESULT 11

ID Q71539 standard; cDNA; 3890 BP.
 AC Q71539;
 DT 28-APR-1995 (first entry)
 DE Brain low Km, cAMP-specific phosphodiesterase cDNA.
 KW Phosphodiesterase; PDE IV(B); brain; ss.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT cds 282..1973
 FT /*tag= a
 FN W09420079-A.
 PD 15-SEP-1994.
 PF 10-MAR-1994; U02612.
 PR 10-MAR-1993; US-029334.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Livi GP, McLaughlin MM, Torphy TJ;
 DR WPI, 94-332667/41.
 DR P-PSDB; R60605.
 PT Human low Km, cAMP-specific phospho-di-esterase and nucleic acids
 PT encoding it - used to identify binding cpds. useful as
 PT therapeutic agents and as probes to evaluate disease states.
 PS Claim 3: Page 29-32; 54pp; English.
 CC The cDNA encodes human low Km, cAMP-specific phosphodiesterase,
 CC PDE IV(B). The enzyme and fusion proteins can be used to identify
 CC ligands and drugs binding to it. These compounds are useful
 CC antidepressant, antialsthmatic and antiinflammatory agents.
 SQ Sequence 3890 BP; 1127 A; 846 C; 852 G; 1065 T;

Query Match 79.0%; Score 15.8; DB 1; Length 3890;
 Best Local Similarity 89.5%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TTTAGGAATCTCCCACTT 20
 Db 2332 TTTATGAATCTTCTCACTT 2350

RESULT 12

ID T34375 standard; cDNA; 3186 BP.
 AC T34375;
 DT 09-OCT-1996 (first entry)
 DE Plasmid pTM72 (ATCC 68602) insert.
 KW Human; glioblastoma cell; heat shock sensitivity; phosphodiesterase;
 KW deficient yeast strain 10DAB; pTM22; rat DPD phosphodiesterase; pdei-;
 KW bovine Ca2+/calmodulin dependent cAMP phosphodiesterase; heart; plasmid;
 KW RAS2(val19); pdei2-; pTM3; pTM72; pRATDPD; pJC99; rolipram sensitive; ss.
 OS Homo sapiens.
 FT Key Location/Qualifiers
 FT cds 139..2349
 FT /*tag= a
 FT /product= cAMP phosphodiesterase
 FN US5527896-A.
 PD 18-JUN-1996.
 PF 20-APR-1990; 511715.

PR 20-APR-1990; US-511715.
 PR 19-APR-1991; US-688352.
 PA (COLD-) COLD SPRING HARBOR LAB.
 PI Colicelli JJ, Wigler MH;
 DR WPI; 96-299902/30.
 DR P-PSDB; W00093.
 PT DNA mols. isolated from human glioblastoma cells - encode
 PT RAS-related or cyclic nucleotide phosphodiesterase proteins
 PS Claim 4; Column 81-88; 101pp; English.
 CC The sequences given in T34374-76 represent plasmid fragments which
 CC contain human glioblastoma cell cDNA inserts which are capable of
 CC correcting the heat shock sensitivity of the phosphodiesterase
 CC deficient yeast strain 10DAB. Several cDNA's were isolated and
 CC sequenced. pTM22 encodes a novel human gene. From computer analysis,
 CC pTM22 putatively encodes a protein homologous to various cAMP
 CC phosphodiesterases, such as the bovine Ca2+/calmodulin dependent cAMP
 CC phosphodiesterase and the rat DPD phosphodiesterase. Sequences related
 CC to pTM22 were found to be expressed in human heart. Plasmid pTM22 was
 CC unable to correct the heat shock sensitivity of RAS2(val19) yeast
 CC strains. It thus appears that the pdei- and pdei2- yeast strain 10DAB
 CC is more sensitive to phenotypic reversion by mammalian cAMP
 CC phosphodiesterase clones than is the RAS2(val19) yeast strain. The
 CC inserts in the plasmids pTM3 and pTM72 were also characterised. These
 CC two different cAMP phosphodiesterase cDNA's were found to be closely
 CC related to, but distinct from, the pRATPD insert and the pJC99 insert.
 CC Biochemical analysis of cell lysates has established that the cDNA's of
 CC pTM3 and pTM72, pJC44x and pRATPD encode rolipram sensitive cAMP
 CC phosphodiesterases.
 SQ Sequence 3186 BP; 927 A; 744 C; 734 G; 781 T;

Query Match 79.0%; Score 15.8; DB 1; Length 3186;
 Best Local Similarity 89.5%; Pred. No. 35;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 TTTAGGAATCTCCCACTT 20
 Db 2706 TTTATGAATCTTCTCACTT 2724

RESULT 13

ID Q28505 standard; DNA; 10798 BP.
 AC Q28505;
 DT 11-JAN-1993 (first entry)
 DE Tomato vacuolar invertase gene (genomic).
 KW RNA polymerase II; promoter; cauliflower mosaic virus; CaMV; ss.
 OS Lycopersicon esculentum.
 FH Key Location/Qualifiers
 FT promoter 1..3519
 FT /*tag= a
 FT /note= "claim 35, page 80"
 FT 2772..3519
 FT /*tag= b
 FT /note= "claim 36, page 80"
 FT 2440..3519
 FT /*tag= c
 FT /note= "claim 37, page 80"
 FT 3471..3474
 FT /*tag= d
 FT 1846..1853
 FT /*tag= e
 FT /note= "sequence matches known plant nuclear
 FT binding sites"
 FT 2012..2019
 FT /*tag= f
 FT /note= "sequence matches known plant nuclear
 FT binding sites"
 FT 2178..2185
 FT /*tag= g
 FT /note= "sequence matches known plant nuclear
 FT binding sites"
 FT 2344..2351
 FT /*tag= h

```

FT      /*tag= h
FT      /note= "sequence matches known plant nuclear
FT      binding sites"
FT      2510..2517
FT      /*tag= i
FT      /note= "sequence matches known plant nuclear
FT      binding sites"
FT      2576..2683
FT      /*tag= j
FT      /note= "sequence matches known plant nuclear
FT      binding sites"
FT      2842..2849
FT      /*tag= k
FT      /note= "sequence matches known plant nuclear
FT      binding sites"
FT      3502
FT      /*tag= l
FT      /label= transcription_start_site
FT      3520..7445
FT      /*tag= m
FT      WO9214831-A.
PN      03-SEP-1992.
PD      21-FEB-1992; U01385.
PF      22-FEB-1991; US-660344.
PR      04-OCT-1991; US-771331.
PA      (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI      Butler WO, Dickinson CD, Elliott KJ, Fitzmaurice LC;
PI      Konno Y, Mirkov TE;
PI      WPI; 92-316183/38.
DR      Tomato vacuolar invertase gene - useful for producing transgenic
PT      plants with higher soluble solid content in fruit, giving greater
PT      commercial value
PS      Disclosure; Page 64-68; 87pp; English.
CC      In order to increase the soluble solids content of tomato fruit
CC      produced by a tomato plant, a DNA construct contg. DNA encoding an
CC      invertase is introduced. The DNA encoding the invertase is pref.
CC      operatively linked to a promoter recognised by the plant RNA
CC      polymerase II. The promoter may be a constitutive promoter,
CC      such as the cauliflower mosaic virus (CaMV) 35S promoter, or a
CC      developmentally regulated promoter that confers fruit specificity
CC      and appropriate temporal control on the expression of the DNA
CC      encoding invertase, e.g. a native Lycopersicon invertase promoter
CC      or the histidine decarboxylase (HDC) promoter region.
CC      If the DNA encodes an invertase that is not a vacuolar invertase,
CC      DNA encoding invertase is operatively linked to DNA that encodes
CC      vacuolar targeting sequences, and, if necessary, DNA encoding
CC      signal sequences.
SQ      Sequence 10798 BP; 3619 A; 1709 C; 1771 G; 3699 T;

Query Match 76.0%; Score 15.2; DB 1; Length 10798;
Best Local Similarity 85.0%; Pred. No. 73;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTTCCCACTT 20
    ||||| ||| ||||| |||||
Db 5874 GTTTATGAAACTTCCAACCTT 5893

RESULT 14
Q28504
ID Q28504 standard; cDNA; 2199 BP.
AC Q28504;
DT 11-JAN-1993 (first entry)
DE Tomato vacuolar invertase gene (cDNA).
KW RNA polymerase II; promoter; cauliflower mosaic virus; CaMV; ss.
OS Lycopersicon esculentum.
FH key Location/Qualifiers
FT cds 7..1917
FT /*tag= a
FT /product= vacuolar_invertase
FT signal_peptide 7..282
FT /*tag= b

```

```

FT      misc_feature 1825..1845
FT      /*tag= c
FT      /label= C-terminal_propeptide
FT      /note= "a sequence effective for vacuolar targeting
FT      comprises at least this sequence"
FT      1798..1914
FT      /*tag= d
FT      /label= C-terminal_propeptide
FT      /note= "vacuolar targeting sequence"
FT      889..927
FT      /*tag= e
FT      /label= active_site
FT      /note= "this region appears to be the active site
FT      of invertase, partic. nucleotides 898..924"
FT      898..924
FT      /*tag= f
FT      /label= active_site
FT      /note= "region 889..927 appears to be the active site
FT      of invertase, partic. nucleotides 898..924"
FT      316..416
FT      /*tag= g
FT      /note= "5' probes" may be derived from this region"
FT      880..980
FT      /*tag= h
FT      /note= "active site" probes may be derived from
FT      this region"
FT      WO9214831-A.
PN      03-SEP-1992.
PD      21-FEB-1992; U01385.
PF      22-FEB-1991; US-660344.
PR      04-OCT-1991; US-771331.
PA      (SALK ) SALK INST BIOTECHNOLOGY IND ASSOC.
PI      Butler WO, Dickinson CD, Elliott KJ, Fitzmaurice LC;
PI      Konno Y, Mirkov TE;
PI      WPI; 92-316183/38.
DR      Tomato vacuolar invertase gene - useful for producing transgenic
PT      plants with higher soluble solid content in fruit, giving greater
PT      commercial value
PS      Claim 7; Page 77 + 61-64; 87pp; English.
CC      In order to increase the soluble solids content of tomato fruit
CC      produced by a tomato plant, a DNA construct contg. DNA encoding an
CC      invertase is introduced. The DNA encoding the invertase is pref.
CC      operatively linked to a promoter recognised by the plant RNA
CC      polymerase II. The promoter may be a constitutive promoter, such
CC      as the cauliflower mosaic virus (CaMV) 35S promoter, or a
CC      developmentally regulated promoter that confers fruit specificity
CC      and appropriate temporal control on the expression of the DNA
CC      encoding invertase, e.g. a native Lycopersicon invertase promoter
CC      or the histidine decarboxylase (HDC) promoter region.
CC      If the DNA encodes an invertase that is not a vacuolar invertase,
CC      DNA encoding invertase is operatively linked to DNA that encodes
CC      vacuolar targeting sequences, and, if necessary, DNA encoding
CC      signal sequences.
SQ      Sequence 2199 BP; 612 A; 465 C; 508 G; 614 T;

Query Match 76.0%; Score 15.2; DB 1; Length 2199;
Best Local Similarity 85.0%; Pred. No. 88;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTTCCCACTT 20
    ||||| ||| ||||| |||||
Db 817 GTTTATGAAACTTCCAACCTT 836

RESULT 15
Q28507
ID Q28507 standard; DNA; 10965 BP.
AC Q28507;
DT 11-JAN-1993 (first entry)
DE Tomato vacuolar invertase gene (genomic).
KW RNA polymerase II; promoter; histidine decarboxylase; HDC;

```


KW cauliflower mosaic virus; CaMV; ss.
OS Lycopersicon pimpinellifolium.
FH Key Location/Qualifiers
FT Promoter 1..3679
FT /*tag= a
FT /note= "claim 34, page 80"
FT tata_signal 3637..3640
FT /*tag= b
FT repeat_unit 1846..1853
FT /*tag= c
FT /note= "sequence matches known plant nuclear
FT binding sites"
FT repeat_unit 2178..2185
FT /*tag= d
FT /note= "sequence matches known plant nuclear
FT binding sites"
FT repeat_unit 2344..2351
FT /*tag= e
FT /note= "sequence matches known plant nuclear
FT binding sites"
FT repeat_unit 2510..2517
FT /*tag= f
FT /note= "sequence matches known plant nuclear
FT binding sites"
FT repeat_unit 2676..2683
FT /*tag= g
FT /note= "sequence matches known plant nuclear
FT binding sites"
FT repeat_unit 2842..2849
FT /*tag= h
FT /note= "sequence matches known plant nuclear
FT binding sites"
FT repeat_unit 3008..3015
FT /*tag= i
FT /note= "sequence matches known plant nuclear
FT binding sites"
FT misc_rna 3668
FT /*tag= j
FT /*label= transcription_start_site
FT precursor_rna 3686..7612
FT /*tag= k
PN WO9214831-A.
PD 03-SEP-1992.
PE 21-FEB-1992; U01385.
PR 22-FEB-1991; US-660344.
PR 04-OCT-1991; US-771331.
PA (SALK) SALK INST BIOTECHNOLOGY IND ASSOC.
PI Butler WO, Dickinson CD, Elliott KJ, Fitzmaurice LC;
PI Konno Y, Mirkov TE;
DR WPI; 92-316183/38.
PT Tomato vacuolar invertase gene - useful for producing transgenic
PT plants with higher soluble solid content in fruit, giving greater
PT commercial value
PS Disclosure; Page 71-75; 87pp; English.
CC In order to increase the soluble solids content of tomato fruit
CC produced by a tomato plant, a DNA construct contg. DNA encoding an
CC invertase is introduced. The DNA encoding the invertase is pref.
CC operatively linked to a promoter recognised by the plant RNA
CC polymerase II. The promoter may be a constitutive promoter,
CC such as the cauliflower mosaic virus (CaMV) 35S promoter, or a
CC developmentally regulated promoter that confers fruit specificity
CC and appropriate temporal control on the expression of the DNA
CC encoding invertase, e.g. a native Lycopersicon invertase promoter
CC or the histidine decarboxylase (HDC) promoter region.
CC If the DNA encodes an invertase that is not a vacuolar invertase,
CC DNA encoding invertase is operatively linked to DNA that encodes
CC vacuolar targeting sequences and, if necessary, DNA encoding signal
CC sequences.
SQ Sequence 10965 BP; 3668 A; 1743 C; 1784 G; 3770 T;

Query Match 76.0%; Score 15.2; DB 1; Length 10965;
Best Local Similarity 85.0%; Pred. No. 73;

Search completed: September 18, 1999, 05:27:20
Job time: 1685 sec

1 GTTTAGGAATCTTCCCACTT 20
||||| ||| ||||| |||||
Db 6041 GTTTATGAACTTCCCACTT 6060

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:34 ; Search time 1405 Seconds
(without alignments)
28.079 Million cell updates/sec

Title: US-09-037-472-4
Perfect score: 20
Sequence: 1 GTTTAGGAATCTTCCCACTT 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:*

- 1: em_est1:*
- 2: em_est2:*
- 3: em_est3:*
- 4: em_est4:*
- 5: em_est5:*
- 6: em_est6:*
- 7: em_est7:*
- 8: em_est8:*
- 9: em_est9:*
- 10: em_est10:*
- 11: em_est11:*
- 12: em_est12:*
- 13: em_est13:*
- 14: em_est14:*
- 15: em_est15:*
- 16: em_est16:*
- 17: em_est17:*
- 18: em_est18:*
- 19: em_est19:*
- 20: gb_est1:*
- 21: gb_est2:*
- 22: gb_est3:*
- 23: gb_est4:*
- 24: gb_est5:*
- 25: gb_est6:*
- 26: gb_est7:*
- 27: gb_est8:*
- 28: gb_est9:*
- 29: gb_est10:*
- 30: gb_est11:*
- 31: gb_est12:*
- 32: gb_est13:*
- 33: gb_est14:*
- 34: gb_est15:*
- 35: gb_est16:*
- 36: gb_est17:*
- 37: gb_est18:*
- 38: gb_est19:*
- 39: gb_est20:*
- 40: gb_est21:*
- 41: gb_est22:*
- 42: gb_est23:*
- 43: gb_est24:*
- 44: gb_est25:*
- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*

- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	16.4	82.0	362	23	H26538 y113g06.r1
2	16	80.0	395	24	H67402 yu52h06.s1
3	16	80.0	298	25	N87127 L2378F Huma
4	15.8	79.0	353	20	T49306 ya74b08.r1
5	15.8	79.0	480	22	H09278 Y198f10.r1
C 6	15.8	79.0	271	22	R41046 Hk464-f Adu
C 7	15.8	79.0	360	22	R47346 He814-r Adu
C 8	15.8	79.0	463	23	H56501 Yt88a03.r1
C 9	15.8	79.0	559	24	H67455 yu52g08.r1
10	15.8	79.0	433	26	W79493 zd78h03.r1
11	15.8	79.0	458	33	AA431115
12	15.8	79.0	377	35	C45344
13	15.8	79.0	376	35	C45344 C45344 YUJ1
14	15.8	79.0	380	35	C48917
15	15.8	79.0	440	38	AA781119
16	15.8	79.0	475	40	AA988380
C 17	15.8	79.0	534	48	C95313
C 18	15.4	77.0	444	22	R48401
C 19	15.4	77.0	374	23	H57709
C 20	15.4	77.0	373	24	H64718
21	15.4	77.0	250	24	H65209
22	15.4	77.0	347	24	H78039
23	15.4	77.0	347	30	AA215780
C 24	15.4	77.0	186	37	AA692500
C 25	15.4	77.0	390	40	AA960919
C 26	15.4	77.0	335	50	AV047898
27	15.2	76.0	592	23	T41588
C 28	15.2	76.0	492	23	D15818
C 29	15.2	76.0	318	23	H32789
C 30	15.2	76.0	549	25	N47378
C 31	15.2	76.0	648	25	N90352
C 32	15.2	76.0	427	26	W33002
C 33	15.2	76.0	441	27	AA045797
C 34	15.2	76.0	230	27	C03156
35	15.2	76.0	478	31	AA290364
C 36	15.2	76.0	515	31	AA307102
C 37	15.2	76.0	430	31	AA310600
C 38	15.2	76.0	340	36	AA651989
39	15.2	76.0	464	37	AA691910
40	15.2	76.0	376	37	AA711473
41	15.2	76.0	113	38	AA737446
42	15.2	76.0	296	38	AA793558
43	15.2	76.0	328	39	F22980
C 44	15.2	5.0	471	40	AA989956
C 45	15.2	76.0	651	54	HS0009377

ALIGNMENTS

RESULT 1
LOCUS H26538 362 bp mRNA
DEFINITION y113g06.r1 Soares breast 2NbHst Homo sapiens cDNA clone
IMAGE:158170 5', mRNA sequence.
ACCESSION H26538
NID G896528
VERSION H26538.1 GI:896528

12-JUL-1995

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevasakis, E., Waterston, R., Williamson, A., Wohldmann, P., and Wilson, R.

TITLE The WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

High quality sequence stops: 203
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: M13RP1
High quality sequence stop: 203.

FEATURES
source
1..362
/organism="Homo sapiens"
/db_xref="GDB:572071"
/db_xref="taxon:9606"
/clone="IMAGE:158170"
/clone_lib="Soares breast 2NBHBst"
/sex="Female"
/dev_stage="adult"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: breast; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCACTGAAGTGGAGCGCGCCCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 66 a 116 c 91 g 82 t 7 others
ORIGIN

Query Match 82.0%; Score 16.4; DB 23; Length 362;
Best Local Similarity 94.4%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 TTTAGGAATCTCCCACT 19
||||||| |||||
Db 129 TTTAGGAACCTCCCACT 146

RESULT 2
LOCUS H67402 395 bp mRNA EST 18-OCT-1995
DEFINITION yu52h06.sl Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:229787 3', mRNA sequence.
ACCESSION H67402
NID g1026142
VERSION H67402.1 GI:1026142
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 395)

AUTHORS Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B., Chisoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W., Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N., Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L., Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J., Trevasakis, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R., and Marra, M.

TITLE Generation and analysis of 280,000 human expressed sequence tags
JOURNAL Genome Res. 6 (9), 807-828 (1996)
MEDLINE 97044478
COMMENT On Jan 24, 1995 this sequence version replaced gi:634117.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Insert Size: 911
High quality sequence stops: 282
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 911 Std Error: 0.00
Seq primer: Promega -21ml3
High quality sequence stop: 282.

FEATURES
source
1..395
/organism="Homo sapiens"
/db_xref="GDB:3780883"
/db_xref="taxon:9606"
/map="21"
/clone="IMAGE:229787"
/clone_lib="Soares fetal liver spleen INFLS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site.1: Pac I; Site.2: Eco RI; 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACGAGAGAAATTAATTAAGATCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 124 a 75 c 100 g 96 t
ORIGIN

Query Match 80.0%; Score 16; DB 24; Length 395;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 TTTAGGAATCTCCCA 17
||||||| |||||
Db 171 TTTAGGAATCTCCCA 186

RESULT 3
LOCUS N87127 298 bp mRNA EST 01-APR-1996
DEFINITION L2378F Human fetal heart, Lambda ZAP Express Homo sapiens cDNA clone L2378 5', similar to REPETITIVE ELEMENT ALU, mRNA sequence.
ACCESSION N87127
NID g1440329
VERSION N87127.1 GI:1440329
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 298)

VERSION H56501.1 GI:1005145
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Euthera; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 463)
Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
Chissoe, S., Dietrich, N., Dubucque, T., Favello, A., Gish, W.,
Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
Trevisan, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
and Marra, M.
Generation and analysis of 280,000 human expressed sequence tags
Genome Res. 6 (9), 807-828 (1996)
97044478
On May 8, 1995 this sequence version replaced gi:801436.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Insert Size: 833
High quality sequence stops: 349
Source: IMAGE Consortium, LLNL
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 833 Std Error: 0.00
Seq primer: M13RPI
High quality sequence stop: 349.

```

FEATURES
source
    Location/Qualifiers
        1..463
            /organism="Homo sapiens"
            /db_xref="GDB:3861331"
            /db_xref="taxon:9606"
            /clone="IMAGE:231340"
            /clone_lib="Soares-pineal_gland_N3HPG"
            /lab_host="DH10B (ampicillin resistant)"
            /note="Organ: pineal gland; Site: PT7T3D (Pharmacia)
            with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
            1st strand cDNA was primed with a Not I oligo(dT) primer
            [5' TGCTTACCATCTCGAGTGGACGGCCGCTTTTTTTTTTTTTTTT
            3'], double-stranded cDNA was size selected, ligated to
            Eco RI adapters (Pharmacia), digested with Not I and
            cloned into the Not I and Eco RI sites of a modified pT7T3
            vector (Pharmacia). Library constructed by Bento Soares
            and M.Fatima Bonaldo."
        94 a 107 C 142 g 115 t 5 others
BASE COUNT
ORIGIN

```

```

Query Match      79.0%  Score 15.8;  DB 23;  Length 463;
Best Local Similarity 89.5%;  Pred. No. 2.3e+02;
Matches 17;  Conservative 0;  Mismatches 2;  Indels 0;  Gaps 0;
QY      2  TTAGGAATCTTCCCACTT 20
DB      396  TTGGGAATCTTCCCCCTT 414

```

RESULT	9	
H67455/c		
LOCUS	H67455	559 bp
DEFINITION	YU59g08_r1 Soares fetal liver spleen INFLS Homo sapiens cDNA clone IMAGE:229790 5', mRNA sequence.	EST 18-OCT-1995
ACCESSION	H67455	
NID	g1026195	
VERSION	H67455.1	GI:1026195
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	
REFERENCE	
AUTHORS	

Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 559)
 Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., DuBuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, J., Lacy, M., Le, N.,
 Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevisan, E., Underwood, K., Wohldmann, P., Waterston, R., Wilson, R.,
 and Marra, M.
 Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)
 97044478
 On Jan 24, 1995 this sequence version replaced gi:6341184.

FEATURES
source

	BASE COUNT	ORIGIN
0	1	0
1	1	1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6
7	1	7
8	1	8
9	1	9
10	1	10
11	1	11
12	1	12
13	1	13
14	1	14
15	1	15
16	1	16
17	1	17
18	1	18
19	1	19
20	1	20
21	1	21
22	1	22
23	1	23
24	1	24
25	1	25
26	1	26
27	1	27
28	1	28
29	1	29
30	1	30
31	1	31
32	1	32
33	1	33
34	1	34
35	1	35
36	1	36
37	1	37
38	1	38
39	1	39
40	1	40
41	1	41
42	1	42
43	1	43
44	1	44
45	1	45
46	1	46
47	1	47
48	1	48
49	1	49
50	1	50
51	1	51
52	1	52
53	1	53
54	1	54
55	1	55
56	1	56
57	1	57
58	1	58
59	1	59
60	1	60
61	1	61
62	1	62
63	1	63
64	1	64
65	1	65
66	1	66
67	1	67
68	1	68
69	1	69
70	1	70
71	1	71
72	1	72
73	1	73
74	1	74
75	1	75
76	1	76
77	1	77
78	1	78
79	1	79
80	1	80
81	1	81
82	1	82
83	1	83
84	1	84
85	1	85
86	1	86
87	1	87
88	1	88
89	1	89
90	1	90
91	1	91
92	1	92
93	1	93
94	1	94
95	1	95
96	1	96
97	1	97
98	1	98
99	1	99
100	1	100

Query Match	2	TT
Best Local		
Matches		384
Qy		TT
Db		TT

RESULT	10
W79493	
LOCUS	
DEFINITION	
ACCESSION	
NID	
VERSION	
KEYWORDS	
SOURCE	
ORGANISM	

ORGANISM	Homo sapiens
ORGANISM	Homo sapiens

REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
The WashU-Merck EST Project
Unpublished (1995)

TITLE
JOURNAL
COMMENT

Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1809 Std Error: 0.00
Seq primer: mob.REGA+ET.

FEATURES

Location/Qualifiers
1. 433
/organism="Homo sapiens"
/db_xref="GDB:1272180"
/db_xref="taxon:9606"
/clone="IMAGE:346805"
/clone_lib="Soares_fetal_heart_NbHL19W"
/sex="unknown"
/dev_stage="19 weeks"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: heart; Vector: pT73D (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGGCCCAATCTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by M.Fatima Bonaldo. This library was constructed from the same fetus as the fetal lung library, Soares fetal lung NbHL19W."

BASE COUNT 122 a 89 c 86 g 136 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 26; Length 433;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTTAGGAATCTTCCCACTT 20
||||| ||||||| |||||
Db 185 TTTATGAATCTTCTCATT 203

RESULT 11
AA431115 458 bp mRNA EST 22-MAY-1997
LOCUS zw9f10.rl Soares_testis_NHT Homo sapiens cDNA clone IMAGE:781483
DEFINITION 5', mRNA sequence.

ACCESSION AA431115
NID 92114823
VERSION AA431115.1 GI:2114823
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS
Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F., Theising, B.,

CONTACT: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-5-6240

White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-Merck EST Project 1997
Unpublished (1997)
On Sep 12, 1996 this sequence version replaced gi:1407164.

Contact: Wilton RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 451.

FEATURES

Location/Qualifiers
1. 458
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:781483"
/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGGCCCAATCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 101 a 127 c 81 g 149 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 33; Length 458;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TTTAGGAATCTTCCCACTT 20
||||| ||||||| |||||
Db 149 TTCAGGAATCTTCCCACTT 167

RESULT 12
C45344 377 bp mRNA EST 10-SEP-1997
LOCUS C45344 Yuji Kohara unpublished cDNA Caenorhabditis elegans cDNA
DEFINITION clone yk393cl 5', mRNA sequence.

ACCESSION C45344
NID 92381597
VERSION C45344.1 GI:2381597
KEYWORDS EST.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans

REFERENCE
AUTHORS
Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis. 1 (bases 1 to 377)
Kohara, Y., Mochizuki, T., Tabara, H., Watanabe, H., Sugimoto, A., Sano, M., Miyata, A. and Nishigaki, A.
Expression map of the C.elegans genome
Unpublished (1996)
On May 9, 1995 this sequence version replaced gi:802599.

Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-5-6240

```

FEATURES
  source
    Email: ykohara@dbj.nig.ac.j.
    Location/Qualifiers
      1..377
        /organism="Caenorhabditis elegans"
        /strain="CBI489 him-8(el489)"
        /db_xref="taxon:6239"
        /clone="yk393cl"
        /clone_lib="Yuji Kohara unpublished cDNA"
        /note="dev_stage-varied, sex=Hermaphrodite male,
        tissue_type=whole animal"
BASE COUNT      102 a 91 c 90 g 91 t 3 others
ORIGIN
Query Match      79.0%; Score 15.8; DB 35; Length 377;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTTCCCACT 19
    ||||| ||||| |||||
Db 11 GTTTATGAATCTTCTCACT 29

RESULT 13
C48917
LOCUS
DEFINITION
  C48917 Yuji Kohara unpublished cDNA Caenorhabditis elegans CDNA
  clone yk450g6 5', mRNA sequence.
ACCESSION
  C48917
NID
  92386179
VERSION
  C48917.1 GI:2386179
KEYWORDS
  EST.
SOURCE
  Caenorhabditis elegans.
  Caenorhabditis elegans.
  Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
  Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
  1 (bases 1 to 376)
  Kohara, Y., Motohashi, T., Tabara, H., Watanabe, H., Sugimoto, A.,
  Sano, M., Miyata, A. and Nishigaki, A.
  Expression map of the C.elegans genome
  Unpublished (1996)
  On Sep 12, 1996 this sequence version replaced gi:1404817.
Contact: Yuji Kohara
Gene Library Lab
National Institute of Genetics
Yata 1111, Mishima, Shizuoka 411, Japan
Tel: 0559-75-0771
Fax: 0559-75-6240
Email: ykohara@dbj.nig.ac.j.
Location/Qualifiers
  1..376
    /organism="Caenorhabditis elegans"
    /strain="CBI489 him-8(el489)"
    /db_xref="taxon:6239"
    /clone="yk524g1"
    /clone_lib="Yuji Kohara unpublished cDNA"
    /note="dev_stage-varied, sex=Hermaphrodite male,
    tissue_type=whole animal"
BASE COUNT      105 a 91 c 91 g 93 t
ORIGIN
Query Match      79.0%; Score 15.8; DB 35; Length 380;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GTTTAGGAATCTTCCCACT 19
    ||||| ||||| |||||
Db 16 GTTTATGAATCTTCTCACT 34

RESULT 15
AA781119
LOCUS
DEFINITION
  a123f08.s1 Soares_testis_NHT Homo sapiens CDNA clone 1391175 3',
  mRNA sequence.
ACCESSION
  AA781119
NID
  92840450
VERSION
  AA781119.1 GI:2840450
KEYWORDS
  EST.
SOURCE
  human.
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
  Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 440)
  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
  Tumor Gene Index
  Unpublished (1997)
  On Sep 12, 1996 this sequence version replaced gi:1400787.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima

```


FEATURES

SOURCE

BASE COUNT
ORIGIN

QY	2	TTTAGGAATCTTCCCACTT	20
Db	61	TGTAGGAATCTTCCCTCTT	79

Search completed: September 18, 1999, 06:47:37
Job time: 3654 sec

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	5	I73227 Sequence 3

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20
|||||
Db 1 TGGCATTGATCTGGTTCATC 20

RESULT 2

HSIL1B 9721 bp DNA 26-JUN-1997
LOCUS Human gene for prointerleukin 1 beta.
DEFINITION X04500
ACCESSION 933788
NID X04500.1 GI:33788
VERSION interleukin 1 beta.
KEYWORDS human.
SOURCE Homo sapiens

ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 9721)
AUTHORS Clark,B.D., Collins,K.L., Gandy,M.S., Webb,A.C. and Auron,P.E.
TITLE Genomic sequence for human prointerleukin 1 beta: possible
evolution from a reverse transcribed prointerleukin 1 alpha gene
JOURNAL Nucleic Acids Res. 14 (20), 7897-7914 (1986)
MEDLINE 87040762
REMARK Erratum: [[published erratum appears in Nucleic Acids Res 1987 Jan
26;15(2):868]]

COMMENT Data kindly reviewed (13-MAY-1988) by Clark B.D.
FEATURES Location/Qualifiers

Source 1..9721
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/clone_lib="leukocyte DNA library"
/map="q13-q24"
669..960
misc_feature
669..960
misc_feature
1576..1581
/note="TATA-box like sequence"
CAAT_signal
1809..1816
CAAT_signal
1859..1866
TATA_signal
1903..1909
Prim.transcript
1934..8953
exon
1934..2005
/number=1
complement(1936..1943)
misc_feature
/note="pot. viral enhancer core sequence"
2006..2465
intron
/number=1
repeat_region
2039..2055
/note="direct repeat 2"
2291..2297
misc_feature
/note="pot. viral enhancer core sequence"
2458..2465
misc_feature
/note="pot. viral enhancer core sequence"
2466..2527
exon
/number=2
Join(2481..2527,3092..3143,5125..5326,5874..6038,
7275..7405,8127..8339)
CDS
/codon_start=1
/product="prointerleukin-1-beta"
/protein_id="CAA28185.1"
/db_xref="PID:g312408"
/db_xref="GI:312408"
/db_xref="SWISS-PROT:P01584"
/translation="MAEVPGLASVMAYVSGNEDDLFEADGPKQMKSFQDLCLPL
DGGIQLRSDHYSGFQRAASVYVAMDKLRKMLVPCQTQFQENDLSTFFPFEEEP
IFFDWNENAVHDPVRLNCTLRDSQKSLVMSGYPYELKALHQQGDMQQVYFSM
SFVQGEENDKIPALGUKENKLYLSVLKDDKPTLQLESVDPKNYPRKMEKRFVN
KIEINNKLEFSAFPNNYISTSAENNPVFLGGTKGGQDITDTMQFVSS"
2528..3091
intron
/number=2
complement(2714..2721)
misc_feature

misc_feature
2858..2865
/note="pot.viral enhancer core sequence"
exon
3092..3143
/number=3
intron
3144..5124
/number=3
misc_feature
4266..4273
/note="pot.viral enhancer core sequence"
misc_feature
4274..4279
/note="pot.viral enhancer core sequence"
misc_feature
4659..4988
/note="Alu repeat"
exon
5125..5326
/number=4
intron
5327..5873
/number=4
misc_feature
5489..5497
/note="pot.viral enhancer core sequence"
exon
5874..6038
/number=5
intron
6039..7274
/number=5
repeat_unit
6205..6212
/note="inverted repeat A"
6213..6220
/note="inverted repeat A"
complement(6910..6915)
misc_feature
/note="pot.viral enhancer core sequence"
complement(6916..6924)
misc_feature
/note="pot.viral enhancer core sequence"
complement(7247..7253)
exon
7275..7405
/number=6
intron
7406..8126
/number=6
misc_feature
7419..7426
/note="pot. viral enhancer core sequence"
exon
8127..8953
/number=7
polyA_signal
8925..8930
/note="pot polyA signal"
8953
polyA_site
9331..9721
misc_feature
/note="Alu repeat"

BASE COUNT 2661 a 2328 c 2122 g 2608 t 2 others
ORIGIN

Query Match 100.0% Score 20; DB 10; Length 9721;
Best Local Similarity 100.0%; Pred.No.1.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20
|||||
Db 1232 TGGCATTGATCTGGTTCATC 1251

RESULT 3

HS283K11 140682 bp DNA PRI 23-APR-1999
LOCUS Human DNA sequence from clone 283K11 on chromosome 6q23.1-24.3.
DEFINITION Contains part of the EVA4 gene for eyes absent (Drosophila) homolog
4. Contains ESTs and GSSs, complete sequence.
ACCESSION AL024497
NID 94680187
VERSION AL024497.5 GI:4680187
KEYWORDS HTG: EVA4; eyes absent.
SOURCE human
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.

```

REFERENCE 1 (bases 1 to 140682)
AUTHORS   Bray-Allen,S.
TITLE     Direct Submission
JOURNAL   Submitted (16-APR-1999) Sanger Centre, Hinxton, Cambridgeshire,
          CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
          requests: clonerequests@sanger.ac.uk
COMMENT   On Apr 25, 1999 this sequence version replaced gi:4581315.
          During sequence assembly data is compared from overlapping clones.
          Where differences are found these are annotated as variations
          together with a note of the overlapping clone name. Note that the
          variation annotation may not be found in the sequence submission
          corresponding to the overlapping clone, as we submit sequences with
          only a small overlap as described above.
          This sequence is the entire insert of clone 283k11. This sequence
          has been finished according to sequence map criteria as follows. An
          attempt is made to resolve all sequencing problems, such as
          compressions and repeats, but not necessarily within known
          annotated human repeat sequence elements (e.g. Alu). Where the
          sequence is ambiguous, there is an annotation using the 'unsure'
          feature key.
          This sequence was generated from part of bacterial clone contigs of
          human chromosome 6, constructed by the Sanger Centre Chromosome 6
          Mapping Group. Further information can be found at
          http://www.sanger.ac.uk/HGP/Chr6
          283k11 is from the library RPC11 constructed at the Roswell Park
          Cancer Institute by the group of Pieter de Jong. For further
          details see http://pacpac.med.buffalo.edu/VECTOR: pcypac2.
FEATURES   Location/Qualifiers
            source          1..140682
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="6"
                        /clone="283k11"
                        /map="q23.1-24.3"
                        /clone_lib="RPC11"
            repeat_region   515..560
                        /note="23 copies 2 mer ag 96% conserved"
            repeat_region   517..560
                        /note="11 copies 4 mer agag 98% conserved"
            repeat_region   704..751
                        /note="MIR repeat: matches 21..70 of consensus"
            repeat_region   1167..1504
                        /note="L2 repeat: matches 2423..2750 of consensus"
            repeat_region   1582..1711
                        /note="L2 repeat: matches 1326..1464 of consensus"
            repeat_region   1729..1995
                        /note="L2 repeat: matches 911..1213 of consensus"
            repeat_region   2289..2459
                        /note="MER58 repeat: matches 1..175 of consensus"
            repeat_region   2479..2630
                        /note="MER5A repeat: matches 9..189 of consensus"
                        complement(2631..3056)
                        /note="match: GSS AQ038582"
            repeat_region   4154..4373
                        /note="MIR repeat: matches 10..247 of consensus"
            repeat_region   5871..5980
                        /note="MIR repeat: matches 77..204 of consensus"
            repeat_region   6726..6904
                        /note="MER5A repeat: matches 2..189 of consensus"
            repeat_region   6929..7114
                        /note="L2 repeat: matches 2054..2276 of consensus"
            repeat_region   7117..7262
                        /note="73 copies 2 mer ta 75% conserved"
            repeat_region   7119..7262
                        /note="36 copies 4 mer tata 75% conserved"
            repeat_region   7127..7246
                        /note="73 copies 40 mer 78% conserved"
            repeat_region   7267..7319
                        /note="L1LMA3 repeat: matches 6252..6304 of consensus"
            repeat_region   7320..7617
                        /note="AluXs repeat: matches 2..299 of consensus"
            repeat_region   7618..7768
                        /note="L1LMA3 repeat: matches 6102..6252 of consensus"
            repeat_region   7769..8046
                        /note="AluJb repeat: matches 3..268 of consensus"
            repeat_region   8047..10633
                        /note="L1LMA3 repeat: matches 3510..6102 of consensus"
            repeat_region   10634..10935
                        /note="AluY repeat: matches 1..301 of consensus"
            repeat_region   10936..15783
                        /note="L1LMA3 repeat: matches 1389..3510 of consensus"
            repeat_region   15785..15967
                        /note="L2 repeat: matches 2263..2440 of consensus"
            repeat_region   15991..16104
                        /note="L2 repeat: matches 2543..2667 of consensus"
            repeat_region   16554..16691
                        /note="MIR repeat: matches 6..151 of consensus"
            repeat_region   17052..17169
                        /note="MIR repeat: matches 20..137 of consensus"
            repeat_region   17182..17217
                        /note="MLT2B repeat: matches 228..265 of consensus"
            repeat_region   18582..18867
                        /note="AluXs repeat: matches 2..286 of consensus"
            repeat_region   18868..18921
                        /note="18 copies 3 mer taa 91% conserved"
            repeat_region   18869..18956
                        /note="44 copies 2 mer aa 68% conserved"
            gene             complement(19490..120213)
                        /gene="EYA4"
            CDS              complement(join(c19490..19573,21006..21142,33874..34039,
                        37747..37826,39841..39984,40128..40270,41425..41491,
                        45957..46049,54426..54494,55851..55975,120164..120213))
                        /gene="EYA4"
            /note="match: proteins Q99504 p97767 Q99502 O08575 O00167
            O60647 P97480; match: cDNAs Y17114 U81602 U81604 U61112
            AF031484 Y17115 U71208 U81603 U61111 AJ000098 AJ000097
            U61110; match: ESTs AAL16744 A1426738"
            /codon_start=1
            /evidence=not_experimental
            /product="DJ283K11.1 (eyes absent (Drosophila) homolog 4)"
            /protein_id="CAB41291.1"
            /db_xref="PID:e1426176"
            /db_xref="PID:g4680188"
            /db_xref="GI:4680188"
            /translation="VKKTCESDVQSQRSMEMQDLASPTLVGGDTGSSKLEK
            SNLSSTVNTGTGENTVLTADLLSCNTSSATMSLLAVTEPLNSETTATTG
            DGALDTGVSITSSGYSRPAHOYSPOLYPSKPYPHILSPAAQTMSAYAGQOYSG
            MOOPAVYATSVTSGQYSLPTYDLGVMPLAKTESGISOQSCLSPGFSYSP
            OPGQTPYSQMPGSSFPAPSTIYANNVSINFGSQDDIPSTAFQNOYAOYISA
            STGAYMTSNTADGTPSTSTYQLOESLPGLTNOPGEFTMQSPFTPKIDERTICR
            SSGSKSRGRKKNPPPPSDLERVFVMDLDETIIVFHSLLTGSYAKYKG"
            repeat_region   21383..21787
                        /note="L2 repeat: matches 872..1276 of consensus"
            repeat_region   21788..22161
                        /note="MER33 repeat: matches 5..323 of consensus"
            repeat_region   22162..22681
                        /note="L2 repeat: matches 1276..1796 of consensus"
            repeat_region   24053..24372
                        /note="L1LME1 repeat: matches 5569..5894 of consensus"
            repeat_region   24444..24728
                        /note="L1LME1 repeat: matches 5270..5560 of consensus"
            repeat_region   24739..24790
                        /note="26 copies 2 mer tg 75% conserved"
            repeat_region   24792..24955
                        /note="MIR repeat: matches 3..192 of consensus"
            repeat_region   25038..25397
                        /note="L1LMA8 repeat: matches 5202..5556 of consensus"
            repeat_region   25398..25769
                        /note="THE1C repeat: matches 1..367 of consensus"
            repeat_region   25770..26466
                        /note="L1LMA8 repeat: matches 5556..6270 of consensus"
            repeat_region   26470..26553
                        /note="FRAM/FAM repeat: matches 5..88 of consensus"
            repeat_region   26646..27252
                        /gene="EYA4"
            misc_feature      /note="match: GSSs AQ342708 AQ169801 AQ423973 B49120"

```

```

misc_feature      complement(27606..27958)
                  /gene="EYA4"
repeat_region     28709..28826 repeat: matches 3..175 of consensus"
repeat_region     29272..29343 repeat: matches 2671..2743 of consensus"
repeat_region     29536..29637 repeat: matches 5195..5300 of consensus"
repeat_region     29636..32712 repeat: matches 2276..5351 of consensus"
repeat_region     32705..33304 repeat: matches 4579..5196 of consensus"
repeat_region     35142..35371 repeat: matches 23..257 of consensus"
repeat_region     35473..36197 repeat: matches 1969..2693 of consensus"
repeat_region     36469..36604 repeat: matches 2571..2710 of consensus"
misc_feature      complement(37165..37585)
                  /gene="EYA4"
                  /note="match: GSS AQ206743"
                  37592..38335
                  /gene="EYA4"
                  /note="match: GSSs AQ122328 AQ103151"
                  38945..39257
                  /note="AluSx repeat: matches 1..312 of consensus"
                  40412..40517
                  /note="MIR repeat: matches 43..147 of consensus"
                  43871..44096
                  /note="MER20 repeat: matches 16..218 of consensus"
                  44105..44151
                  /note="MER3 repeat: matches 1..47 of consensus"
                  44152..44231
                  /note="MER3 repeat: matches 123..209 of consensus"
                  44358..44417
                  /note="MER3 repeat: matches 142..207 of consensus"
                  47478..47692
                  /note="MIR repeat: matches 20..261 of consensus"
                  47957..48165
                  /note="MIR repeat: matches 8..237 of consensus"
                  49367..49695
misc_feature      Query Match      92.0%; Score 18.4; DB 10; Length 140682;
                  Best Local Similarity 95.0%; Pred. No. 21;
                  Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGCATTGATCTGGTTCATC 20
    |||||
Db 122831 TGGCATTGATCTGGTACATC 122850

RESULT 4
LOCUS      XLU36194 1069 bp mRNA VRT 12-OCT-1995
DEFINITION Xenopus laevis helix-loop-helix transcription factor hairy1 mRNA,
complete cds.
ACCESSION  U36194
NID         91017760
VERSION     U36194.1 GI:1017760
KEYWORDS    African clawed frog.
SOURCE      Xenopus laevis
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Amphibia;
            Batrachia; Anura; Mesobatrachia; Pipidae; Xenopodinae;
            Xenopus.
REFERENCE   1 (bases 1 to 1069)
AUTHORS     Dawson,S.R., Turner,D.L., Weintraub,H. and Parkhurst,S.M.
TITLE       Specificity for the Hairy/Enhancer of split bHLH proteins maps
            outside the bHLH domain, and suggests two separable modes of
            transcriptional repression
JOURNAL     Mol. Cell. Biol. (1995) In press
REFERENCE   2 (bases 1 to 1069)

AUTHORS     Turner,D.L.
TITLE       Direct Submission
JOURNAL     Submitted (12-SEP-1995) David L. Turner, Genetics, Fred Hutchinson
            Cancer Research Center, 1124 Columbia St., A3-025, Seattle, WA
            98104, USA
FEATURES     Location/Qualifiers
             1..1069
             /organism="Xenopus laevis"
             /db_xref="taxon:8355"
             27..827
             /codon_start=1
             /function="helix-loop-helix transcription factor"
             /product="hairy1"
             /protein_id="AAA79185.1"
             /db_xref="PID:g1017761"
             /db_xref="GI:1017761"
             /translation="MPADVMEKNSSSPVAATPASVSNTPDKPKTASEHRKSSKPIMEK
             RRARINESLQGLTKLILDAKKDSRHSKLEKADILEMTVKHLNLRQVMSAALST
             DSVLGKRYAGFSECHNEVTRFLSTCEGVNTEVTRLLHLNCHNQTQINAMTYPGQPH
             PSAALPHPAYQPMVOLPGAPOSSGPVIACTMGGPPVEAAKYGGFQLVPADPGGFA
             FLITNPAPFHNGSVIPVTNSVNGTALPSPSVSPSVTIDSVWRPW"
BASE COUNT      284 a 281 c 253 g 251 t
ORIGIN           1
Query Match      87.0%; Score 17.4; DB 4; Length 1069;
Best Local Similarity 94.7%; Pred. No. 28;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TGGCATTGATCTGGTTCAT 19
    |||||
Db 483 TGGCATTGATCTGGTTCAT 465

RESULT 5
LOCUS      HUMHAIRY/c 5552 bp DNA PRI 21-APR-1995
DEFINITION Human HRY gene, complete cds.
ACCESSION  L19314
NID         9436999
VERSION     L19314.1 GI:436999
KEYWORDS    Homo sapiens DNA.
SOURCE      Homo sapiens
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 5552)
AUTHORS     Feder,J.N., Li,L., Jan,L.Y. and Jan,Y.N.
TITLE       Genomic cloning and chromosomal localization of HRY, the human
            homolog to the Drosophila segmentation gene, hairy
JOURNAL     Genomics 20 (1), 56-61 (1994)
MEDLINE     94292187
FEATURES     Location/Qualifiers
             1..5552
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /map="3q28-q29"
             /chromosome="3"
             join(1707..1814,1943..2038,2287..2373,3006..3557)
             /gene="HRY"
             join(1707..1814,1943..2038,2287..2373,3006..3557)
             /gene="HRY"
             /note="also called HHL"
             /codon_start=1
             /protein_id="AAA65220.1"
             /db_xref="PID:g780128"
             /db_xref="GI:780128"
             /translation="MPADIMEKNSSSPVAATPASVNTTPDKPKTASEHRKSSKPIMEK
             RRARINESLQGLTKLILDAKKDSRHSKLEKADILEMTVKHLNLRQVMSAALST
             DPSVLGKRYAGFSECHNEVTRFLSTCEGVNTEVTRLLHLNCHNQTQINAMTYPGQPH
             PALQAPPPPPGPGPOHAFPPPLVPIPGGAAPPPGAPCKLGSQAGAAKYFGG
             FOVVPADPGGFAFLIPNGFAHSGPVPVVTNSGTSVGNVSPSSGSPSLTADSMWR
             PWRN"

```

```
BASE COUNT      1322 a      1548 c      1410 g      1272 t
ORIGIN

Query Match      87.0%; Score 17.4; DB 9; Length 5552;
Best Local Similarity 94.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
|||||
Db 3171 TGGCATTGATCTGGTTCAT 3153

RESULT 6
LOCUS      HSPC13M      5037 bp      mRNA      PRI      23-NOV-1995
DEFINITION      H.sapiens encoding PC1/PC3.
ACCESSION      X64810 S88573
NID      g35317
VERSION      X64810.1 GI:35317
KEYWORDS      endoprotease; PC1/PC3 protein; subtilisin homologue.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS      Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5037)
Creemers,J.W., Roebroek,A.J. and Van de Ven,W.J.
Expression in human lung tumor cells of the proprotein processing
enzyme PC1/PC3. Cloning and primary sequence of a 5 kb cDNA
FEBS Lett. 300 (1), 82-88 (1992)
JOURNAL      92192290
MEDLINE
AUTHORS
REFERENCE      2 (bases 1 to 5037)
Roebroek,A.J.M.
Direct Submission
Submitted (12-JUN-1992) A.J.M. Roebroek, Universitaire
Ziekenhuizen, Leuven, Centrum voor Menselijke Erfelijkheid, UZ
Gasthuisberg, Herestraat 49 3000 Leuven, BELGIUM
JOURNAL
FEATURES
source
1. .5037
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="adult"
/tissue_type="human lung cancer: carcinoma tumor"
5'UTR
1. .189
/citation=[1]
190. .2451
/function="proprotein processing enzyme"
/citation=[1]
/codon_start=1
/product="PC1/PC3"
/protein_id="CA446031.1"
/db_xref="PID:g35318"
/db_xref="GI:35318"
/transalt="MERRAHSOCTAFVLCFAMCALNSAKAKROFVNEWAAETPGGPE
AASAIAEELGDLGQISLEHLYFKHNHPRSRSAFHITKRLSDDDRVIWAEQ
YEKRSKRSLRDSALNLPNDPMWNOQLDTRWTAALPKDLHVPVWQKITGK
VVTLDLGGVHNDIYANDPEASIDFNNDHDPFPRYDNIENKHGTRCAGEIAM
QANNHKGGVAYNSKVGIRMLDGIIVDAIEASSIGFNPGHVYDSAGWNDGKT
VQGPGLAQAFEGYKQGRGSGIFVWASNGRQGNCDGTYDSIYTSISSA
SQGLSPWAEKCSLTATSYGGDYTDQITTSADLHNDCTETHTGTSASAPLAGIF
ALALEANPLTWDMOHLVWTSYDPLANNPWKNGAGLWNSRSGFGLNNAKALY
DIADPTWRSVPKECVKNDPEPRALKANGEVILEIPTRACGEQNAIKSLRHVQ
FEATLEYSRGLHVLTSNAGTSTVLLAERDTPNGFKNDFMSVHTGENPIGT
WTLRTDMSGRIGNEGRVNWKLILHGTSSQPEHMKOPRYVTSYNTVNDRRGVEKM
DPGEQPTQENKENTLVKSPSSSSVGGRRDELEGAPQAMRLQLQSAFKSNPPK
QSPKSPSAKLNIPIYENFYEALKEKLNKPSQLKDSDSLNDYVDYFNTKPKYKRRDR
LQALVDILNEEN"
2449. .5037
1458 a      1061 c      1106 g      1412 t
BASE COUNT
ORIGIN

Query Match      87.0%; Score 17.4; DB 9; Length 5552;
Best Local Similarity 94.7%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
|||||
Db 3171 TGGCATTGATCTGGTTCAT 3153

RESULT 6
LOCUS      HSPC13M      5037 bp      mRNA      PRI      23-NOV-1995
DEFINITION      H.sapiens encoding PC1/PC3.
ACCESSION      X64810 S88573
NID      g35317
VERSION      X64810.1 GI:35317
KEYWORDS      endoprotease; PC1/PC3 protein; subtilisin homologue.
SOURCE      human.
ORGANISM      Homo sapiens
REFERENCE      Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS      Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5037)
Creemers,J.W., Roebroek,A.J. and Van de Ven,W.J.
Expression in human lung tumor cells of the proprotein processing
enzyme PC1/PC3. Cloning and primary sequence of a 5 kb cDNA
FEBS Lett. 300 (1), 82-88 (1992)
JOURNAL      92192290
MEDLINE
AUTHORS
REFERENCE      2 (bases 1 to 5037)
Roebroek,A.J.M.
Direct Submission
Submitted (12-JUN-1992) A.J.M. Roebroek, Universitaire
Ziekenhuizen, Leuven, Centrum voor Menselijke Erfelijkheid, UZ
Gasthuisberg, Herestraat 49 3000 Leuven, BELGIUM
JOURNAL
FEATURES
source
1. .5037
/organism="Homo sapiens"
/db_xref="taxon:9606"
/dev_stage="adult"
/tissue_type="human lung cancer: carcinoma tumor"
5'UTR
1. .189
/citation=[1]
190. .2451
/function="proprotein processing enzyme"
/citation=[1]
/codon_start=1
/product="PC1/PC3"
/protein_id="CA446031.1"
/db_xref="PID:g35318"
/db_xref="GI:35318"
/transalt="MERRAHSOCTAFVLCFAMCALNSAKAKROFVNEWAAETPGGPE
AASAIAEELGDLGQISLEHLYFKHNHPRSRSAFHITKRLSDDDRVIWAEQ
YEKRSKRSLRDSALNLPNDPMWNOQLDTRWTAALPKDLHVPVWQKITGK
VVTLDLGGVHNDIYANDPEASIDFNNDHDPFPRYDNIENKHGTRCAGEIAM
QANNHKGGVAYNSKVGIRMLDGIIVDAIEASSIGFNPGHVYDSAGWNDGKT
VQGPGLAQAFEGYKQGRGSGIFVWASNGRQGNCDGTYDSIYTSISSA
SQGLSPWAEKCSLTATSYGGDYTDQITTSADLHNDCTETHTGTSASAPLAGIF
ALALEANPLTWDMOHLVWTSYDPLANNPWKNGAGLWNSRSGFGLNNAKALY
DIADPTWRSVPKECVKNDPEPRALKANGEVILEIPTRACGEQNAIKSLRHVQ
FEATLEYSRGLHVLTSNAGTSTVLLAERDTPNGFKNDFMSVHTGENPIGT
WTLRTDMSGRIGNEGRVNWKLILHGTSSQPEHMKOPRYVTSYNTVNDRRGVEKM
DPGEQPTQENKENTLVKSPSSSSVGGRRDELEGAPQAMRLQLQSAFKSNPPK
QSPKSPSAKLNIPIYENFYEALKEKLNKPSQLKDSDSLNDYVDYFNTKPKYKRRDR
LQALVDILNEEN"
2449. .5037
1458 a      1061 c      1106 g      1412 t
BASE COUNT
ORIGIN
```

```
Query Match      87.0%; Score 17.4; DB 10; Length 5037;
Best Local Similarity 94.7%; Pred. No. 38;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
|||||
Db 4122 TGGCATTGAACTGGTTCAT 4140

RESULT 7
LOCUS      AE000841/c      10051 bp      DNA      BCT      16-APR-1998
DEFINITION      Methanobacterium thermoautotrophicum from bases 525786 to 535836
ACCESSION      AE000841 AE000666
NID      g2621665
VERSION      AE000841.1 GI:2621665
KEYWORDS      Methanobacterium thermoautotrophicum.
SOURCE      Methanobacterium thermoautotrophicum
ORGANISM      Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
Methanobacterium.
REFERENCE      1 (bases 1 to 10051)
AUTHORS      Smith,D.R., Doucette-Stamm,L.A., Deloughery,C., Lee,H.-M.,
Dubois,J., Aldredge,T., Bashirzadeh,R., Blakely,D., Cook,R.,
Gilbert,K., Harrison,D., Hoang,L., Keagle,P., Lumm,W., Pothier,B.,
Qiu,D., Spadafora,R., Vicare,R., Wang,X., Wierzbowski,J.,
Gibson,R., Jiwani,N., Caruso,A., Bush,D., Safer,H., Patwell,D.,
Prabhakar,S., McDougall,S., Shimer,G., Goyal,A., Pietrowski,S.,
Church,G.M., Daniels,C.J., Mao,J.-I., Rice,P., Nolling,J. and
Reeve,J.N.
Complete genome sequence of Methanobacterium thermoautotrophicum
deltaH: functional analysis and comparative genomics
J. Bacteriol. 179 (22), 7135-7155 (1997)
98037514
REFERENCE      2 (bases 1 to 10051)
AUTHORS      Smith,D.R.
TITLE      Direct Submission
JOURNAL      Submitted (10-AUG-1997) Genomics and Technology Development, Genome
Therapeutics Corporation, 100 Beaver Street, Waltham, MA
02154-8448, USA
FEATURES
source
1. .10051
/organism="Methanobacterium thermoautotrophicum"
/strain="delta H"
/db_xref="taxon:2166"
/clone="MTH"
162. .2117
/gene="MTH587"
162. .2117
/gene="MTH587"
/note="Function Code:10.07 - Metabolism of Macromolecules,
Aminoacyl tRNA synthetases and tRNA modification; similar
to, sp:LN:SYM_METJA AC:Q58659, p(-)5.3E-154, pid=468"
/codon_start=1
/transl_table=11
/product="methionyl-tRNA synthetase"
/protein_id="AAB85093.1"
/db_xref="PID:g2621666"
/db_xref="GI:2621666"
/transalt="MSKVFTICALPYANGPTHLGLRSTYIPADIYARYRRLRGDDVL
FYCATDEHTPIAVKAEAGRTPLEVATRYHEMIRGDLERCDSFDSFTRTTDELHHE
IAQKFFKLYEKGFYKEDIKOLYCGDQORFLPDRVEGTCYCGSEGARGDCECG
RILEPLQLEDPRCMVCGSEPEYVDSRHVFLNQODEIRENIEGSEFNVNRYAIQ
WLRGLKWLTRDMGWQVPLEGNEGRIITVWGAEFLGYISSAAASKKTKGRDWR
YDSSAIFIGKDIILYHAFIPALLMAYGCTRPANIAGELYSLGQKMSKNVVV
WTSDFERDRLRYLYTNAPLRTDTSWDDFQRRVNDLADVLGFLHRTFSFT
GRFDRVPAAGELTAEDLEFLSRAAHDVAELDKRFOFDALMHIILAKRGKY
FNDQEPWAKVESPASTCLHCLNLAANGKLLRPSPSAAGRLISMNLEDAWG
FHELREGQVIERAKPLFSKITDEAIEEKAKLIEEDDEVEITVDFTIMDIRGVIR
SAERIEGSDKLLKLLIDVGERMQVVAELAEKISPEDLVVERKITVLVNLKAPLFGVR
SEGAVLATGESLNILDPGDAEVGERIR"
2155. .3732
gene
```

```
/gene="MTH588"
/2155..3732
/gene="MTH588"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pir:LN:E64438 AC:E64438, p()-1.6E-14, pid=19%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85094.1"
/db_xref="PID:g2621667"
/db_xref="GI:2621667"
/translation="MNEGLTARSPFLSUKDKVSLDDLSKVEGFIDLYLRGNL
ELQDLKAMELRGKYPFRISGYSSQYSEIADVEDHDIKHAQYFMKASTKNLL
DRVNSAISHRILNLEIYALQPCDSRLSRLESEYEDIDLDGVMPCGSGRMEI
FSGSAICPEIIPHLPSGDYVMKMSLSMARKAFKIMRLFKNEKKGAVRSAITLVI
KLYEDGRIRRTITDSDNDYERMLRGYGNVRIELIOFHRKKSIIINDRYTRTAL
ALGYAGLSDIIMDKERYSEKLDROYDAVKNRILFEARTSPDLRSLSEDLQEV
VOKMDHLHEAGLGGSGRLNRELREDLEVVEVKRELFRDVPYNLMDVALYLG
SLDRSKANGPPNLRPLVDRSQVTFDELSRDVLLNTCTGEGMYIDKLDVLLK
KEIEDKMKGLHMKPNLAFGAALHMEGEADIETCAEIFHVRVEEVKRNIEHLG
KPSDRKXMLNLKIGD"
/3725..4201
/gene="MTH589"
/3725..4201
/gene="MTH589"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pir:LN:E64484 AC:B64484, p()-8.9E-10, pid=28%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85095.1"
/db_xref="PID:g2621668"
/db_xref="GI:2621668"
/translation="MIDPMAEKIHVNQPLTSRIELLEKNPDLEKEITCPSLHYRHS
RYLDALDELGVNVPKRGIPRYKTNDVKLVQSLKEGKTPKQISGIIINPLKTV
YYLKGDIKLRGKRRKYDRNTRLRVREARNMGMPARKISKDLGILRTVYVYLKNG"
/4323..5321
/gene="MTH590"
/4323..5321
/gene="MTH590"
/note="Function Code:11.02 - Cell envelope, Surface
polysaccharides and lipopolysaccharides; similar to,
pir:LN:H64438 AC:H64438, p()-4.4E-40, pid=27%"
/codon_start=1
/transl_table=11
/product="N-acetylglucosamine-1-phosphate transferase"
/protein_id="AAB85096.1"
/db_xref="PID:g2621669"
/db_xref="GI:2621669"
/translation="MPDYVYGDSMVLKLTLEYSSVLLVSLICCIIFTSTVMPRLI
NKLKANVYGNDIHKISKPIVAEMGGIGILFGFTIGMFGMYCFPELOYELMVTLLVI
LLVGIVGVDDILVLSRREKLELLFLAGLPIIWAAPPKVGILYHIMMFAVSIASNLT
NMLAGLIESGLSIAMTALFASCIIMKGYDVSIIITWMLGALLAFMLNRYPSRVF
PGDVGTLIIGACIASVAFIGRVKIIIALIVLNNIIDGLKFKYSAGVVERHHRPTEIA
EDGKLIAPPEGNSLIRWLRPMTKEKVMVITWVSGMFFGLVGLFIPLDPF"
/complement(5333..6220)
/gene="MTH591"
/complement(5333..6220)
/gene="MTH591"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, gp:GI:g1001351, p()-0.000061, pid=72%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85097.1"
/db_xref="PID:g2621670"
/db_xref="GI:2621670"
/translation="MMMFIVAVLGFIDFLRGEDEGESLEKYLKRLKLEEDFTLIVRI
AGFYAEGTREALDYLERAYVASEMDELMFALDSMGDVYLSDRKVTAMFYFM
EALRIYTSNPLRADLREKIKEVEKIREAMDIAINRLRESESGMPEVDLEVEPL
LNLRYDSVOSLTLYEAGSYEDSIOIMEAYOIAIDGDTSEASQLLIGLYSLKNE
FDERSRYLKAKALFRKTNNELGLVALVLLGLTDFIRNADGVASSFRAVEILQKL
DEKLESVTLLENLTLSF"
complement(6362..6408)
/note="47 bp direct repeat includes part of intergenic
sequence; 100% ID to interval 532261-532215"
complement(6430..6476)
/note="47 bp direct repeat includes part of MTH592
(conserved protein - paralog.fam. 32); 100% ID to interval
532193-532147"
complement(6446..7120)
/gene="MTH592"
complement(6446..7120)
/gene="MTH592"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pir:LN:F64469 AC:F64469, p()-0.00000016,
pid=25%"
/transl_table=11
/product="conserved protein"
/protein_id="AAB85098.1"
/db_xref="PID:g2621671"
/db_xref="GI:2621671"
/translation="MSENYSTGIAGLDAVLGKPEPGVYLMLLTGPPKVGKTVLATEFTY
RALQRRNPCLFIAAEYGYRDLQRNTAFNNFIOQLADKLHVIDLPTGLAGKPRDSGN
VRYSAIQNTTDLVKVGISTRSLYQEGTMSVFDASII:AFNPPKLVRLVKAYND
RVRDAGGIALVITYTAGVADEIEEIPGLFDVHIQMDGSVISAMDGERTAPYMTD
QGKVG"
complement(7122..7823)
/gene="MTH593"
complement(7122..7823)
/gene="MTH593"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, pir:LN:F64469 AC:F64469, p()-4.2E-12, pid=28%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85099.1"
/db_xref="PID:g2621672"
/db_xref="GI:2621672"
/translation="MIVRIFSGIRGLDELGGASIPENTVTVYVGPVKVKSIFAYGF
AAGGAEMDEPLYISADYGLSDLRNMAVLGMDPEYSLQKELLYVIDAISISGPGD
EGSTYSSVHNPTDIMVKLGVAIRNITNRYSMFRSLDSTLTLMFNDMLIVRLT
AYIMRIKEAGGTAVVITYEGSADPKVETMLKAVVDNIHLDSGSELTVEAMVGTGRSA
PYTTDDSGIKVWHEG"
/complement(7824..9005)
/gene="MTH594"
complement(7824..9005)
/gene="MTH594"
/note="Function Code:14.01 - Unknown, Conserved protein;
similar to, gp:GI:e198926:g1103582, p()-7.2E-10, pid=12%"
/codon_start=1
/transl_table=11
/product="conserved protein"
/protein_id="AAB85100.1"
/db_xref="PID:g2621673"
/db_xref="GI:2621673"
/translation="MNNKTYIPKLDLLGGVKEGASIMFASPGVDYBAFGYOMLN
GRLDEGTRGRIFTNVPEPESIIIEFRSYGNDIESHENGLEFVDGSSPELGMPSDQR
YTVNDYSEIKDVLNAINDIPGGVGINNLSVIDYLCNGDTLEIVANNRASSELVD
NLYVTFEMDEYEPEDIDKRESMCDVNNLTIEERVYIQQGFWMVANSWSPETMVL
FFVQPGGKAVYVPEKILVPTGMSKGSFVATKTSVDRKALSAPFTTIANDIGH
LEYKGFMDVIFGTQGFDMILDLVLSREAVGATILIDSTAPETFAKEMIRKTRAF
AIPRVVANKQDLPALSPERIREMMKLGADVPVIVPASVTEGWSGVRDALDTLLGLLYG
G"
complement(9002..9382)
/gene="MTH595"
complement(9002..9382)
```

Query Match

Best Local Similarity 84.0%; Score 16.8; DB 1; Length 10051;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGCATTGATCTGTTTCATC 20

||||| | ||||| |||

Db 6912 TGGCATTCAACTGGTTCATC 6893

```
RESULT 8
GOGDH/c 2890 bp DNA BCT 15-MAR-1999
LOCUS G. oxydans adh gene.
DEFINITION G. oxydans adh gene.
ACCESSION X62710.560040
NID 958416
VERSION X62710.1 GI:58416
KEYWORDS coenzyme PQQ; gdh gene; gluconic acid production; glucose
dehydrogenase; quinoprotein.
SOURCE Gluconobacter oxydans.
ORGANISM Gluconobacter oxydans.
Bacteria; Proteobacteria; alpha subdivision; Acetobacteraceae;
Gluconobacter.
REFERENCE 1 (bases 1 to 2890)
AUTHORS Cleton-Jansen,A.M., Dekker,S., van de Putte,P. and Goosen,N.
TITLE A single amino acid substitution changes the substrate specificity
of quinoprotein glucose dehydrogenase in Gluconobacter oxydans
JOURNAL Mol. Gen. Genet. 229 (2), 206-212 (1991)
MEDLINE 92017653
REFERENCE 2 (bases 1 to 2890)
AUTHORS Goosen,N.
TITLE Direct Submission
JOURNAL Submitted (07-FEB-1992) N. Goosen, Dept of Mol Genetics, Gorlaeus
Laboratories, Leiden University, P O Box 9502, 2300 RA Leiden, THE
NETHERLANDS
FEATURES
source
Location/Qualifiers
1..2890
/organism="Gluconobacter oxydans"
/db_xref="taxon:442"
242..245
RBS
gene
251..2677
/gene="gdh"
251..2677
CDS
/gene="gdh"
/EC_number="1.1.99.17"
/codon_start=1
/transl_table=11
/product="glucose dehydrogenase
(pyrrroloquinoline-quinone)"
/protein_id="CAA44594.1"
/db_xref="PID:g4493391"
/db_xref="GI:4493391"
/db_xref="SWISS-PROT:P27175"
/translation="MSTISRLRPLMALITAAVFAALCGAILTVGGAWVAIGGLPVYVIL
GLALLATFLSFRNPAAIYLFVAVVGVTVIWEITVVGDIWALIPRSDIIVILGIWL
LIPFVSRAVSARGPSCRSFVAVGLFASLFTDPDHDISGELPTDIAASPADPD
NVPASEWYAGRTQAGDRWSPLNQINATVSNLKVAMHHTKDMWNSNDPGEATNEAT
PIEFNTLYMGLSHOKLFAVDGATGNVKKVYDPKLOINPGQHILTCRGVSHETPANA
MDSGNAPATDCAKDSILPVNDGRLEVDADTGKTCGFGNNGEIDLRVPNQYTPPG
QYEPSPVITDKLIIANSAITDNGSVKQASGATQAFDVTYTKRVWVFDSNPDNPQL
PDESHVPFHPNSNWSIVSSYDANLNLVYIPMGVTPDQMGDRTKDSERFAPGIVL
NADTKLAWFYQTVHHDLMDLPSQSLVSDTQKDTLPAIYAPTKGTGDI FVLDR
TGKETVAPETVPQGAAPGHTSPTQPMSTLTLPKPNLNDSDIMGTFIDMPCSI
VFTLURYGEPPTPSSKGLIFPGDLGMFNGGLAVDPQRYVANPISLPFVSQLPV
RPGNPLPEENAKGTGETGLQNYGIPYAVNLFPDLPVLLPFGKMPORPTPWGY
VAGDILKTKVYVQHRNGLTRDMSMTGSLPIPLPIKIGVPSLGGPLSTAGNLGLTA
SMDIYIRAYNTTGVKLVMDRLPAGAAQTPIYAINKGQIYTVYAGGHNSSPTRMGDD
IIAYALPQK"
misc_feature 293..664
note="membrane spanning fragment"
mutation 2612
/gene="gdh"
/note="mutation changing substrate specificity"
BASE COUNT 555 a 975 c 786 g 574 t
ORIGIN
Query Match 84.0%; Score 16.8; DB 1; Length 2890;
Best Local Similarity 90.0%; Pred. No. 71;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 1 TGCATTGATCTGGTTCATC 20
| | | | | | | | | | | | | | | |
Db 812 TCGCATTGATCTGGTTCAGC 793

RESULT 9
ATY12227/c
LOCUS ATY12227
DEFINITION Arabidopsis thaliana DNA, 24 kb surrounding PFL locus.
ACCESSION Y12227
NID 92505864
VERSION Y12227.1 GI:2505864
KEYWORDS GTP-binding protein; hypothetical protein; poly(A) binding protein;
ribosomal protein S18A; RPS18A gene; topoisomerase.
SOURCE thale cress.
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
Rosidae; Caprales; Brassicaceae; Arabidopsis.
REFERENCE 1 (bases 1 to 24053)
AUTHORS Terryn,N.
TITLE Direct Submission
JOURNAL Submitted (02-APR-1997) N. Terryn, University Gent, Lab of
Genetics, VIB, K.L. Ledeganckstraat 35, 9000 Gent, BELGIUM
REFERENCE 2 (bases 1 to 24053)
AUTHORS Terryn,N., Neyt,P., De Clercq,R., De Keyser,A., Van Den Daele,H.,
Ardiles,W., Dehais,P., Rouze,P., Gielens,J., Villarroel,R. and Van
Montagu,M.
TITLE Sequence analysis of a 24-kb contiguous genomic region at the
Arabidopsis thaliana PFL locus on chromosome 1
MEDLINE 98034136
COMMENT Related sequences: Y09821, Y09822, Y09823, Z23165, Z28701.
FEATURES
source
Location/Qualifiers
1..24053
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/map="30 +/- cm (at PFL locus)"
/clone="cc17J13"
<1..1857
/note="gene1"
/codon_start=1
/product="putative topoisomerase"
/protein_id="CAA72903.1"
/db_xref="PID:e1173395"
/db_xref="PID:g2505865"
/db_xref="GI:2505865"
/db_xref="SPTREMBL:O64378"
/translation="DPTSCFDADLSEYKKKATVIVEYFGTNDVYVSVNELKELGMAE
YRIFVKKLVSMAMDRHDKEMAAFLSTLYADVIDPPEYVIRGNKLVASADDLSVD
IPDAVGLAVFAVAIVDDILPPALFKQKMLPDNSKGVLEVKAKESYATPLHAE
VPEKRWGTONTAEDVKARINDLLKEVMSGDKKEAFRCIKGLKVPFFHETVLRAL
IMAKERRKAVRLDLKETIEVGLINSTQVTKGFSRIIDSLDIDIPARRILQS
FISKAASEGVLKASSLSADAGEKLEKSNANFVKAKAKSIITREIFSGDTSEVSH
CLTELNASSQLRAIFVKVLTILAMDRKKREKEMACVLSTLGFPPKVSFAFSLI
ESADDTALDNPVVVEDLAMEFLAVDVDEVLAPDLLEVLNQTPGKVSFAFSLI
TLKARLSGERILKRGGGGIEITNSPGSTVEKVEKIQILIEEYVGGDLREASRCVK
ELMGPFHVEVKKSVVRIIEKENEEERLWLLKVCDFDSGLVTIYQMTGPKRYDESJ
EDLSLDVDDAAKKFSTALKKESLKGSTNHSLLRIRKVRNRTARLLPDDNI"
<2325..2377
/number=1
join(<2325..2377,2471..2497,2618..2717,2816..2962,
3059..3139,3223..3368,3515..>3572)
/note="gene2"
join(2325..2377,2471..2497,2618..2717,2816..2962,
3059..3139,3223..3368,3515..3572)
/note="gene2; homologous to EST T76449; corresponding
cDNA: Y09821"
/codon_start=1
/product="GTP-binding protein Rab7"
```


AC007067 89934 bp DNA PLN 17-JUN-1999
Genomic sequence for Arabidopsis thaliana BAC T10024 from
Chromosome 1, complete sequence.

AC007067
NID 9455821
KEYWORDS
SOURCE HTG.
ORGANISM Arabidopsis thaliana
thale cress.

REFERENCE
AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Khan, S., Kim, C., Walker, M., Altafi, H., Araujo, R., Conn, L., Conway, A. B., Gonzalez, A., Hansen, N. F., Huizar, L., Kremenetskaia, I., Lenz, C., Li, J., Liu, S., Luross, S., Rowley, D., Schwartz, J., Toriumi, M., Vysotskaia, V., Yu, G., Davis, R., W., Federspiel, N., Theologis, A., and Ecker, J. R.
TITLE Genomic sequence for Arabidopsis thaliana BAC T10024 from Chromosome 1
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 89934)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 89934)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 89934)
AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Khan, S., Kim, C., Walker, M., Altafi, H., Araujo, R., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Huizar, L., Kremenetskaia, I., Lenz, C., Li, J., Liu, S., Luross, S., Rowley, D., Schwartz, J., Toriumi, M., Vysotskaia, V., Yu, G., Davis, R., Federspiel, N., Theologis, A., and Ecker, J.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT On Apr 2, 1999 this sequence version replaced gi:4464282.

FEATURES
Location/Qualifiers
1..89934
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(120..350)
/feature="unknown: similar to EST gb|N96021"
/codon_start=1
/evidence=not_experimental
/product="t10024.1"
/protein_id="AAD39561.1"
/db_xref="PID:g5091532"
/db_xref="GI:5091532"
/translation="MGCVSPKCVGAKLRSSKRRKRRRIQKRAVSSRLSDG
SFDNLDDHHRNFPSSRGLSLRSLVSSAFD"
complement(join(5929..6152,6393..6466,6659..6726,
6855..6932,7008..7297,7466..7502))
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="t10024.2"
/protein_id="AAD39562.1"
/db_xref="PID:g5091533"
/db_xref="GI:5091533"
/translation="MDHVSYKVFNIENRALBEKLCHETELQSKDNIISLAOLE
SEKSNHQHVEIISLSELCATFASKLVGHFSLICLLDQYLOKTLQVKKDAVE
NLISEKALYSENKGMILQIQESVSLTEEDRKVFTSILTFEAGSDERNKRSHN

AC007067 89934 bp DNA PLN 17-JUN-1999
Genomic sequence for Arabidopsis thaliana BAC T10024 from
Chromosome 1, complete sequence.

AC007067
NID 9455821
KEYWORDS
SOURCE HTG.
ORGANISM Arabidopsis thaliana
thale cress.

REFERENCE
AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Khan, S., Kim, C., Walker, M., Altafi, H., Araujo, R., Conn, L., Conway, A. B., Gonzalez, A., Hansen, N. F., Huizar, L., Kremenetskaia, I., Lenz, C., Li, J., Liu, S., Luross, S., Rowley, D., Schwartz, J., Toriumi, M., Vysotskaia, V., Yu, G., Davis, R., W., Federspiel, N., Theologis, A., and Ecker, J. R.
TITLE Genomic sequence for Arabidopsis thaliana BAC T10024 from Chromosome 1
JOURNAL Unpublished (1999)
REFERENCE 2 (bases 1 to 89934)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (12-MAR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 3 (bases 1 to 89934)
AUTHORS Ecker, J. R.
TITLE Direct Submission
JOURNAL Submitted (01-APR-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th Street and Hamilton Walk, Philadelphia, Pennsylvania 19104-6018, USA
REFERENCE 4 (bases 1 to 89934)
AUTHORS Shinn, P., Brooks, S., Buehler, E., Chao, Q., Dunn, P., Khan, S., Kim, C., Walker, M., Altafi, H., Araujo, R., Conn, L., Conway, A., Gonzalez, A., Hansen, N., Huizar, L., Kremenetskaia, I., Lenz, C., Li, J., Liu, S., Luross, S., Rowley, D., Schwartz, J., Toriumi, M., Vysotskaia, V., Yu, G., Davis, R., Federspiel, N., Theologis, A., and Ecker, J.
TITLE Direct Submission
JOURNAL Submitted (17-JUN-1999) Arabidopsis thaliana Genome Center, Department of Biology, University of Pennsylvania, 38th and Hamilton Walk, Philadelphia, PA 19104-6018, USA
COMMENT On Apr 2, 1999 this sequence version replaced gi:4464282.

FEATURES
Location/Qualifiers
1..89934
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="1"
complement(120..350)
/feature="unknown: similar to EST gb|N96021"
/codon_start=1
/evidence=not_experimental
/product="t10024.1"
/protein_id="AAD39561.1"
/db_xref="PID:g5091532"
/db_xref="GI:5091532"
/translation="MGCVSPKCVGAKLRSSKRRKRRRIQKRAVSSRLSDG
SFDNLDDHHRNFPSSRGLSLRSLVSSAFD"
complement(join(5929..6152,6393..6466,6659..6726,
6855..6932,7008..7297,7466..7502))
/feature="hypothetical protein"
/codon_start=1
/evidence=not_experimental
/product="t10024.2"
/protein_id="AAD39562.1"
/db_xref="PID:g5091533"
/db_xref="GI:5091533"
/translation="MDHVSYKVFNIENRALBEKLCHETELQSKDNIISLAOLE
SEKSNHQHVEIISLSELCATFASKLVGHFSLICLLDQYLOKTLQVKKDAVE
NLISEKALYSENKGMILQIQESVSLTEEDRKVFTSILTFEAGSDERNKRSHN

CDS

AKVLTGDIHGLFNDKLTREVKFKHFKSQSITMHRAGTTSTQKNVLOKK"
complement(20664..21317)
/note="similar to germin-like protein (gi|2935521);
similar to ESTs gb|T88481 and gb|A1099566"
/codon_start=1
/evidence=not_experimental
/product="t10024.7"
/protein_id="AAD39567.1"
/db_xref="PID:g5091538"
/db_xref="GI:5091538"
/translation="MILNLLTLTLLMRVSRDPPDLQDYCVSPPPSHQQLFNGKL
CKDPTQASVDFSTALSNGTKTKPMINVTITANLPGNLVGLTMRDLFGGS
GVYPPHPRASEVTCQIDVLLGVFTSGRFTQELHPGETFVFKGLIHFLXNID
TVSSALVSGLSQNGTQIVSLSFISFKPPFLVFLKSAVDINGQDVARIKSLG"
complement(23710..24085,24185..24255,24338..24415,
24503..24579,24668..24845))
/note="similar to response regulator 1 (dbj|BA311431);
similar to ESTs gb|T43772, emb|Z34204, gb|AA067391, and
emb|Z34620"
/codon_start=1
/evidence=not_experimental
/product="t10024.8"
/protein_id="AAD39568.1"
/db_xref="PID:g5091539"
/db_xref="GI:5091539"
/translation="MARGGVCLRRSEMSVGGIGGIESAPLDLDEVHLAVDDSLV
DRIVLERLITTSCKVTADSGRALEFLGDNKESAERFORLKLVDLITDYMFGMT
GYELKKIKESNREVPVIMSSNWLTRDCLREGAQDFLLAPVKLAKRURSH
LTKDVKLNGKRLKIDESSVNSLPPSPPLTIPSSPPPLTVSTESSDSSPLSP
VEFTSPSPSPIDDDDLTSSSESPIRRQMRSPGLD"
30820..31455
/note="putative zinc finger protein 5 (pir|S55885)"
/codon_start=1
/evidence=not_experimental
/product="t10024.9"
/protein_id="AAD39569.1"
/db_xref="PID:g5091540"
/db_xref="GI:5091540"
/translation="MSINPTMSRTGESSSGSSDKTIKFGFELISGRTPEITAES
VSSNTTSLTMKRHECYGCFANSSQALGGHQAHKRLKKRQLQAGRNATG
YLTNHOQPIITSQRYKTSYCAFSSMHVNDQMGVINDSSRSQIINRRNDTC
QDLNQSQGMKLVGRPNMIOFQRLSSRSDQMSINSDLHLGFGAGDA"
Join(34675..34779,34868..34959,35199..35370,35452..35574,
35654..35827,35940..36050,36130..36263,36349..36442,
36641..36739,36824..36975,37035..37180,37669..37792,
37887..38001,38055..38154,38234..38286,38510..38686,
38770..38807,38875..38968,39046..39225,39319..39433,
39530..39599,39729..39876,40074..40202,40333..40458,
40612..40677,40772..41011)
/note="unknown; similar to ESTs dbj|C26360, gb|T45760,
and gb|AA394518"
/codon_start=1
/evidence=not_experimental
/product="t10024.10"
/protein_id="AAD39570.1"

Query Match 84.0%; Score 16.8; DB 8; Length 89934;
Best Local Similarity 90.0%; Pred. No. 1.4e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATGATCGTTCATC 20
|||||||
Db 60429 TGGCATGATCGATTC 60448

RESULT 11
T22J18
LOCUS 83163 bp DNA PLN 26-AUG-1998
DEFINITION Arabidopsis thaliana chromosome 1 BAC T22J18 sequence, complete
ACCESSION AC003979
NID 93172156
VERSION AC003979.1 GI:3172156

HTG.
SOURCE
ORGANISM
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
1 (bases 1 to 83163)
Vysotskaia,V.S., Schwartz,J.R., Toriumi,M., Yu,G., Oji,O., Kwan,A.,
Liu,S., Li,J., Araujo,R., Au,M., Brendel,V., Buehler,E.,
Conway,A.B., Conway,A.R., Dewar,K., Feng,J., Kim,C., Kurtz,D.,
Li,Y., Palm,C.J., Shinn,P., Sun,H., Davis,R.W., Ecker,J.R.,
Federspiel,N.A. and Theologis,A.
Arabidopsis thaliana chromosome 1 BAC T22J18 sequence, complete
sequence
Unpublished (1998)
This sequence is of BAC T22J18 from Arabidopsis thaliana chromosome
1. The sequence does not represent the sequence of the entire
insert of this clone. It is shorter by 11541 bp because we submit
only the unique sequence of the clone. However, in order to
facilitate the joining of overlapping clones in the future for
creation of larger contigs, we provide a small overlap (200 bp)
between overlapping submitted clones. The 5' end of this sequence
overlaps by 200 bp the 3' end of the sequence of the BAC F19G10.
2 (bases 1 to 83163)
Theologis,A.
Direct Submission
Submitted (07-JAN-1998) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
3 (bases 1 to 83163)
Theologis,A.
Direct Submission
Submitted (02-JUN-1998) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
4 (bases 1 to 83163)
Theologis.
Direct Submission
Submitted (03-JUL-1998) Plant Gene Expression Center, 800 Buchanan
St., Albany, CA 94710, USA
On Jun 2, 1998 this sequence version replaced gi:2887489.
The sequence of BAC T22J18 from Arabidopsis thaliana chromosome 1.
FEATURES
Location/Qualifiers
1..83163
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
complement(<68..1372)
/gene="T22J18.1"
complement(<68..1372)
/gene="T22J18.1"
/note="Contains similarity to membrane associated
salt-inducible protein homolog IG002N01.30 gb|2191151 from
A. thaliana BAC gb|AF007269. EST gb|F14461 comes from
this gene. Gene continues on the 3' end of BAC F19G10
gb|AF000657 gene F19G10.21."
/codon_start=1
/evidence=not_experimental
/protein_id="AAC3504.1"
/db_xref="PID:g3287676"
/db_xref="GI:3287676"
/translation="MPSSPSRSLRLTVSEICKFIPQSWKOLPRPISKTHDEES
VPQVLFNSFRHCISGQIYEAFTFSLRYSOSGHEFVLYSSALLSTCVGFNEVPV
QQLHACISSGLEFSDVLVPLKLVTFYSAFNLLDEAQTITENSEILHPLPMVNLGYSI
RNKRFQSVSYVKRMKSGIRADEFTYPSVFKACALIDFAYGRVHGSIEVSHRCN
LYCNALISMTYKRFQVDVRLDFRMSRDVSNNAIINCYTSEEKLGFAFKLLDRM
YLSGVEASIVTWNTIAGCLEPAGNYIGALNCVVGMRNCNFIQSVAMINGKACSHIG
ALKWGVFHCILVIRSCFSFSDIDNVRNSLITWYSCDRLHAFIVFOQVEANSUSTWN
SIISGPAYNERSEETSFLLKMLLSGFPNHTLASILPLFAR"
3324..3758
/gene="T22J18.2"
3324..3758
/note="T22J18.2"
/gene="T22J18.2"
/note="Contains similarity to transcription factor (TINY)
isolog T02004.22 gb|2062174 from A. thaliana BAC

TITLE Conservation and diversity of eukaryotic translation initiation factor eIF3

J. Biol. Chem. 272 (2), 1101-1109 (1997)

79150873

2 (bases 1 to 2898)

Asano, K. and Hershey, J.W.B.

Direct Submission

Submitted (12-JAN-1996) Katsura Asano, Biological Chemistry,

University of California at Davis, School of Medicine, Building

MS1A, Davis, CA 95616, USA

Location/Qualifiers

1..2898

/organism="Homo sapiens"

/db_xref="taxon:9606"

/cell_line="HeLa"

50..2791

/codon_start=1

/product="translation initiation factor eIF-3 p110 subunit"

/protein_id="A003462.1"

/db_xref="PID:g1718197"

/db_xref="GI:1718197"

/translation="MSRFTTGSSESSLSGEEELVTKPVGGNYGKQPLLSEDEED

TKRVRSADKRFELTNLIRIRNAKIRDTVKCLEFELGKAYGKAKSIVDKGV

PRYIRILADLEDYLNELWDEKGMKNNKNAKALSTLRQKIRKYNRDFESHITSYK

ONPEQSADEAKNEEDSGSDEDEGVSATFLKKSAPSGSRFLKKWDE

DESDSDDEDDWGTSSPDSSEEGKOTALASFLKKAFTDEKKAERKRED

KAKKHDRKRLDEEEDNDEGGWVRGVPVLPKPKFAGTEITHAVVIKLN

EILQARGKGTDRAGLQVLQIAANNLGEVIVIKFNFIASLYDYNPLATY

MKPEMGKCLDCINELMDILFANPNIFVGENILESENHNADQPLRVGCGILFLVER

MDEFTKIMONTDPSQVEVHLKDEAOVCAIIRVORYLEKGTTEVCRIYLRL

HTYKFDYKAHQRLTPPGSSKSEQDAENEGDSAVLMERLCKYIVAKDRTDRI

CAILCHIVHALLSRWYOARDLMLSHLQDNIOHADPPVQIILYNTWVLGICAFROG

LTKAHNALLDIQSGRAKELLGQGLLRQENQOEKERRRQVFFHLHINLELL

ECVYLSAMLLLEIPYMAAHESDARRRMISKQFHHQLRVGERQPLLPSPESMRHVAA

SKAMKGMKTCFIFINERKNGKVDLPFPAKRVRLMLVKIQEESLYLFTSYV

YDSIMETLSDFELDTVHSIIISKMIINEELMASLDQPTQTVVMHRTPTAQONLA

LQAEKGLSVENNERVFDHKQGTGYGVRDQKDGKRGYMRGGYRQOQSQTAY"

2898

/note="47 A nucleotides"

/polya_site

BASE COUNT 772 a 762 c 819 g 545 t

ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 2898;

Best Local Similarity 90.0%; Pred. No. 71;

Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 TGGCATTGATCTGTTTCATC 20

||||||| ||| |||||

Db 455 TGGCATTGTTCTGTTTCATC 436

||||||| ||| |||||

RESULT 13

HUAC002544

LOCUS

DEFINITION Homo sapiens Chromosome 16 BAC clone C11987/SK-A-761H5, complete

sequence.

ACCESSION AC002544

NID 93337382

VERSION AC002544.1

KEYWORDS GI:3337382

SOURCE HTG.

ORGANISM human.

Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 156641)

Adams, M.D., Loftus, B.J., Zhou, L., Crosby, M., Fuhrmann, J.,

Mason, T.M., Brandon, R., Kim, U.J., Kerlavage, A.R. and Venter, J.C.

Homo sapiens Chromosome 16 BAC clone C11987/SK-A-761H5

unpublished

2 (bases 1 to 156641)

Adams, M.D. and Loftus, B.J.

Direct Submission

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Submitted (12-SEP-1997) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA, Email:

mdadams@tigr.org

3 (bases 1 to 156641)

Adams, M.D. and Loftus, B.J.

Direct Submission

Submitted (24-JUL-1998) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

On Jul 24, 1998 this sequence version replaced gi:2959563.

Address all correspondence to: Mark Adams The Institute for Genomic

Research 9712 Medical Center Dr, Rockville, MD 20850, USA e-mail

address: humgen@tigr.org. The orientation of the sequence is from

SP6 end to 3' end. Genes were identified by a combination of five

methods including: XGRAIL (available by anonymous ftp from

arthur.epm.ornl.gov), Genefinder (Phil Green, University of

Washington), Genscan (Chris Burge,

http://gnomic.stanford.edu/~chris/GENSCANW.html) searches of the

complete sequence against a peptide database, and the Human gene

Index database at tigr (http://www.tigr.org/tdb/hgi/hgi.html).

Genes without peptide homology having spliced EST hits are termed

'Unknown gene product'. Genes encoding tRNAs are predicted by

trnscan-SE (Sean Eddy, http://genome.wustl.edu/eddy/trnscan-SE/).

Location/Qualifiers

1..156641

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="16"

/map="16p11.2"

/clone="A-761H5"

7426..7531

/note="44331, SHOC-36038, Chr. 16, Homo sapiens"

/db_xref="dbSTS:G30050"

27016..108426

/gene="A-761H5.3"

/pseudo

27383..27567

/gene="A-761H5.3"

/note="25260, CHLC.GCT16G06, Chr. -, Homo sapiens"

/db_xref="dbSTS:G16020"

complement(join(32539..32821,32917..33079,33182..33269,

33358..33423,33522..33681,33998..36361,40957..41181,

41407..41597,43190..43283,43419..43544,43739..43880,

44649..44752,44877..45036,52772..52899,53278..53387,

53516..53570,53760..53885,54108..54297,54440..54505,

56420..56550,56689..56743))

/gene="A-761H5.1"

complement(32539..56743)

/gene="A-761H5.1"

complement(join(32739..32821,32917..33079,33182..33269,

33358..33423,33522..33681,33998..36361,40957..41181,

41407..41597,43190..43283,43419..43544,43739..43880,

44649..44752,44877..45036,52772..52899,53278..53387,

53516..53570,53760..53885,54108..54297,54440..54505,

56420..56550,56689..56743)

/gene="A-761H5.1"

/codon_start=1

/product="Translation initiation factor eIF-p110"

/protein_id="AAC27426.1"

/db_xref="PID:g3337383"

/db_xref="GI:3337383"

/translation="MSRFTTGSSESSLSGEEELVTKPVGGNYGKQPLLSEDEED

TKRVRSADKRFELTNLIRIRNAKIRDTVKCLEFELGKAYGKAKSIVDKGV

PRYIRILADLEDYLNELWDEKGMKNNKNAKALSTLRQKIRKYNRDFESHITSYK

ONPEQSADEAKNEEDSGSDEDEGVSATFLKKSAPSGSRFLKKWDE

DESDSDDEDDWGTSSPDSSEEGKOTALASFLKKAFTDEKKAERKRED

KAKKHDRKRLDEEEDNDEGGWVRGVPVLPKPKFAGTEITHAVVIKLN

EILQARGKGTDRAGLQVLQIAANNLGEVIVIKFNFIASLYDYNPLATY

MKPEMGKCLDCINELMDILFANPNIFVGENILESENHNADQPLRVGCGILFLVER

MDEFTKIMONTDPSQVEVHLKDEAOVCAIIRVORYLEKGTTEVCRIYLRL

HTYKFDYKAHQRLTPPGSSKSEQDAENEGDSAVLMERLCKYIVAKDRTDRI

CAILCHIVHALLSRWYOARDLMLSHLQDNIOHADPPVQIILYNTWVLGICAFROG

LTKAHNALLDIQSGRAKELLGQGLLRQENQOEKERRRQVFFHLHINLELL

ECVYLSAMLLLEIPYMAAHESDARRRMISKQFHHQLRVGERQPLLPSPESMRHVAA

```
SKAMKMDGKTCISFIINERKMGKVWDLFPPEADKVRVTRLMLVRKIQEESLRTYLFTYSV
YDSISMTLSDFMLDPTVHSIISKMIINEELMASLDQPTOTVVMHRTPTAQNLA
LOLAELKGLSVENNERVEDHKQGTGYGFRDQDKGYRKNEGYMRGGRQOQSOTAY"
57091..57208
/gene="A-761H5.3"
/notes="7608, STS1-CSRL-24g1-uA/CSRL-24g1-uz, Chr. -, Homo
sapiens"
/db_xref="dbSTS:G02122"
/join(<66573..66995,67730..>68008)
/gene="A-761H5.2"
<66573..>68008
/gene="A-761H5.2"
/join(66573..66995,67730..68008)
/gene="A-761H5.2"
/codon_start=1
/product="CDC37-like gene"
/protein_id="AAC27427.1"
/db_xref="PID:g3337384"
/db_xref="GI:3337384"
/translation="MALGGSTSLRTTVQGVSWAEAKCWTTCAGATLSDDERHFN
THRALFRWOHARVELMEQFOKEKELDRCWEYCKMAECQKLEKLEYSVGARVE
VWLHELSNEQLEENPKKSTWNVDTLKSKSFRUGEMDPTKAKFPQOOCIDAGLR
VPNSEAEKEGEECSGTCWKLFPPQAGMSVRDPPQLQLPPVCKPLGAPFOKTKID
TISPAGFLHFRCSF"
complement(join(<109422..110066,111426..111461,
111571..111631,115712..115907,116020..116119,
123403..>123468))
/gene="A-761H5.4"
complement(<109422..>123468)
/gene="A-761H5.4"
complement(join(109422..110066,111426..111461,
111571..111631,115712..115907,116020..116119,
123403..123468))
/gene="A-761H5.4"
/codon_start=1
/product="Unknown gene product splice form-1"
/protein_id="AAC27428.1"
/db_xref="PID:g3337385"
/db_xref="GI:3337385"
/translation="MLRFLWLLIWLGLFTSHOPTVPSFLKTFWRSRNGHDSMDVQQ
RAWRNSKQGLRSICMHTKRVSSFRGNKIGLKDVITLRRHVETKVRXKIRKRVV
TKINRDKTKRKQKQFORQELRRRAEDYHKCKIPPSARKPLCNCWVRMAAE
HRSSGILPWLYLTATLKNRMRQPPPTQOHSITDNSLSLKTTPPECILTLPPSVD
DNIEKCPAPLPSPSPSVDNDLKECLFVLPSPPLPSVDNLKECLFVLPSPPL
PVSVDNLKTPPLATQEAKEVEKPPKRWVRDVEGSPKRRVRDVEGSPAPKRRQ
EAAQQLPKRRRLSKLRHCTQAWAIRNP"
109425..109537
/gene="A-761H5.4"
/notes="40991, SHGC-32146, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G27136"
complement(join(110411..111461,111571..111631,
115712..115907,116020..116119,123403..>123468))
/gene="A-761H5.5"
complement(110411..>123468)
/gene="A-761H5.5"
complement(join(111219..111461,111571..111631,
115712..115907,116020..116119,123403..123468))
/gene="A-761H5.5"
/codon_start=1
/product="Unknown gene product splice form-2"
/protein_id="AAC27429.1"
/db_xref="PID:g3337386"
/db_xref="GI:3337386"
/translation="MLRFLWLLIWLGLFTSHOPTVPSFLKTFWRSRNGHDSMDVQQ
RAWRNSKQGLRSICMHTKRVSSFRGNKIGLKDVITLRRHVETKVRXKIRKRVV
TKINRDKTKRKQKQFORQELRRRAEDYHKCKIPPSARKPLCNCWVSLFVFL
AFEHSLPGQMDTFFSLQCAQLQREMAERKAAAYRHHSPIPVGNRVQKHLHPVPG
PLT"
121375..121480
/gene="A-761H5.5"
/notes="44231, SHGC-36058, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G30050"
complement(join(<130544..130663,130765..130905,
```

```
135131..135224,135353..135408,135503..135571,
135652..135698,137020..137132,139362..139505,
139593..139665,140471..140556,140677..140756,
141606..141677,142305..142401,144492..144570,
144724..144966))
/gene="A-761H5.6"
complement(<130544..144946)
/gene="A-761H5.6"
complement(join(130544..130663,130765..130905,
135131..135224,135353..135408,135503..135571,
135652..135698,137020..137132,139362..139505,
139593..139665,140471..140556,140677..140756,
141606..141877,142305..142401,144492..144570,
144724..144769))
/gene="A-761H5.6"
/codon_start=1
/product="CLN"
/protein_id="AAC27430.1"
/db_xref="PID:g3337387"
/db_xref="GI:3337387"
/translation="MGCAGSRRFSDSEGETVPEPRLPLDHOAGKNAVGFWLL
GLCNFSYVVMLSAAHDLILPSPRVLVSGICAGSFVAFSHSVCTSLCGVVFASI
ADILPTLVILAPLGLHLLPYSPRVLVSGICAGSFVAFSHSVCTSLCGVVFASI
SSGLGEVTELSLTAIFYPRAVISWSSGTGGAGLGLSYLGLQAGLSPOOTLLSMLG
IPALLASFLLLTSPAQDPGGEAEASARQPLIRTEAPESKPGSSSLSRERWT
VFKGLWYIVPLVVVYFAEYFINOGLFELLFFWNTSLSHAQYRWYQMLYQAGVFAS
SSLRCCIRFTWALIQCLNVLADWVGFELPSYLVFLIILIEGLLGGAAVYVNT
FHNIALETSDHEHREFAATCISDTLIGLSLGLALPLHDFLCQLS"
BASE COUNT 41602 a 38891 c 38211 g 37878 t 59 others
ORIGIN
```

```
Query Match 84.0%; Score 16.8; DB 11; Length 156641;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGCATTGATCTGTTTCATC 20
||||| ||| ||| ||| |||
Db 53837 TGCATTGTTCTTGTTCATC 53856

RESULT 14
HU091326
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
HUU091326 150296 bp DNA PRI 22-OCT-1998
Human Chromosome 16 BAC clone CIT987SK-A-234F9, complete sequence.
U91326
G1931583
U91326.1 GI:1931583
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 150296)
Adams,M.D., Loftus,B.J., Phillips,C.A. and Venter,J.C.
Direct Submission
Submitted (04-MAR-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
2 (bases 1 to 150296)
Adams,M.D., Loftus,B.J., Zhou,L., Venter,J.C., Phillips,C.,
Brandon,R., Fuhrmann,J. and Kim,U.J.
Direct Submission
Submitted (10-APR-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, Maryland 20850, USA
3 (bases 1 to 150296)
Adams,M.D.
Direct Submission
Submitted (28-JUL-1998) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
BAC clone CIT987SK-234F9 is located in band 16p11.2 o
f chromosome 16. Genes were identified by a combination of five
methods: XGRAIL (available by ano
nymous ftp from arthur.epm.ornl.gov), Genefinder (available by
```

anonymous ftp from colineu.washington.edu), GENSCAN (available using the e-mail server at gencan@genomic.stanford.edu), searches of the EST database at TIGR (<http://www.tigr.org/tdb/hcd/hcd.html>) and searches against a peptide database. Repeats were identified using RepeatMasker (Smit, A.F.A. and Green, P. unpublished, <http://ftp.genome.washington.edu/rm/RepeatMasker.html>).

FEATURES

source
1..150296
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/map="16p11.2"
/clone="A-234F9"
1826..2518
/gene="A-234F9.1"
/pseudo
43758..43870
/note="40991, SHGC-32146, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G27136"
43758..44232
/db_xref="dbSTS:T71136"
complement(43984..64312)
/gene="A-234F9.3"
/pseudo
complement(55458..55769)
/gene="A-234F9.3"
/db_xref="dbSTS:"
55657..55762
/gene="A-234F9.3"
/note="44231, SHGC-36058, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G30050"
75320..75594
/db_xref="dbSTS:G16020"
complement(80543..81004,81158..81245,81334..81399,81498..81657,83981..84344,88929..89153,89379..89569,91162..91255,91391..91516,91711..91852,92624..92721,92846..93011,100749..100876,101254..101363,101492..101546,101736..101861,102084..102273,102416..102481,104395..104525,104664..104718)
/gene="A-234F9.2"
complement(80543..104718)
/gene="A-234F9.2"
complement(80664..80746,80842..81004,81158..81245,81334..81399,81498..81657,83981..84344,88929..89153,89379..89569,91162..91255,91391..91516,91711..91852,92624..92721,92846..93011,100749..100876,101254..101363,101492..101546,101736..101861,102084..102273,102416..102481,104395..104495)
/gene="A-234F9.2"
/codon_start=1
/product="eIF-3 p110 subunit"
/protein_id="AAC27674.1"
/db_xref="PID:g1931584"
/db_xref="GI:1931584"
/translation="MSRFTTGSSESSLSGELVTPVCGNGTKQPLLSEDEEDTKRVRSKDKREELTNLIRINAMKIROVTKLEEFELGKAYGKAKSLVDKEGVPYRIILADLYELNEDWEGKKMKNNAKSLTRQKIRYNRDFESHITSYKQNPQSADEAKNEDESGSDEDEDEGSAATFLKKSEAPSGSRKFLKRWDDDEDESDDEDMDTGSTSDSDEEGKQTALASRLFKAPTDDKAAEKKRDKAKKHDKRSKRLDEEDNEGGEWVRGVPLVKVPKMFAGTEITHAVVKKLLEILQAKGKGTDRAAQIELLOLLVQIAENNLGEGVIVIKFNITASLYDNPNTATYMKPEMKGKCLDCINELDLFANFNIFVGENILESENLDHNDQPLRVGCLITLVERMDEFTKIMQNTDPHSEVYVHLKDEAQCAIIEVORYLEEGKTEEVLYLIRILHTYKTDYKAHQRLTPQSSKSEQDAENEGDSAVLMERLCKYIYAKDRDRIRTCAILCHYHALLSRVQARDLMLSHLDNQHDAPPVQIILNRTMVLGICAFRQGLTKDAHALLDIQSSRAKELGGLLRSLOERNOEQEKVRRRQVPPFLHNLLELLKCYLVLSAMLELIPYWAHESDARRWISKQFHOLRVGERQLPGPSPREHYVAAEKAMKQDMLTCHSFINKNMKGWDLFPKADKVRMLVVKRIQESLRIYLYTYSVYDSISMETLSDMFELDPTVHSTISKMIINEELMSLDQPTOTVVMHRTPEFAQONLAQLAEKLSGVLENNEVRFDHKQGTGYGFRDQKGYRKNKGVMRRGGYRQQQSQTAY105065..105182

STS

/note="7608, STS1-CSRL-2491-ua/CSRL-2491-u2, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G02122"
145648..145778
/note="24052, SHGC-17356, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G15633"
146724..146836
/note="40991, SHGC-32146, Chr. 16, Homo sapiens"
/db_xref="dbSTS:G27136"
39918 a 36953 c 36142 g 37133 t 150 others
ORIGIN

Query Match 84.0%; Score 16.8; DB 11; Length 150296;
Best Local Similarity 90.0%; Pred. No. 1.6e+02;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20
||||||| ||| |||||
Db 101813 TGGCATTGTTCTTGTTCATC 101832

RESULT 15

MLCB33/c
LOCUS MLCB33 42224 bp DNA BCT 24-JUN-1997
DEFINITION Mycobacterium leprae cosmid B33.
ACCESSION 294723
NID 92052097
VERSION 294723.1 GI:2052097
KEYWORDS acetyltransferase; cdp-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase; Coiled coil protein; glycosyl transferase; metE; methionine synthase; pgsA; ppk; pyruvate phosphate dikinase; recA; REPLEP; sporulation protein.

SOURCE

ORGANISM Mycobacterium leprae
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacteriaceae; Mycobacteriaceae;
Mycobacterium.
1 (bases 1 to 42224)
AUTHORS Badcock, K. and Churcher, C.M.
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 42224)
2 (bases 1 to 42224)
Barrell, B.G. and Rajandream, M.A.
Direct Submission
Submitted (21-APR-1997) Mycobacterium leprae sequencing project, Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Dr. Stewart T. Cole, [3] Unite de Genetique Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France Requests for cosmids should be sent to Karin Eiglmeier (kei@pasteur.fr)

REFERENCE

AUTHORS Eiglmeier, K., Honore, N., Woods, S.A., Caudron, B. and Cole, S.T.
TITLE Use of an ordered cosmid library to deduce the genomic organization of Mycobacterium leprae
JOURNAL Mol. Microbiol. 7 (2), 197-206 (1993)
MEDLINE 93188700
COMMENT

Notes:
The Sanger Centre is funded to complete the sequence of M. leprae by the Heiser Program for Research in Leprosy and Tuberculosis of The New York Community Trust.
Work in Paris is supported by the Heiser Trust, the Association Francaise Raoul Follereau and the Groupement de Recherches et des Etudes des Genomes (GIF-GEIG).
Details of M. leprae sequencing at the Sanger Centre are available on the World Wide Web.
(URL, <http://www.sanger.ac.uk/pathogens/>)
CDS are numbered using the following system eg MLCB33.01c. ML (M. leprae), CB33 (cosmid name), .01 (first CDS), c (complementary strand).
The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. All CDS over 100 codons have been analysed. Gene prediction is based on positional base preference in codons especially where there is an increase in the observed/expected third position G + C. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, or ttg) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid #CONTIG #LOCATION.

FEATURES

```

source
  1. 42224
    /organism="Mycobacterium leprae"
    /db_xref="taxon:1769"
    /clone="cosmid B33"
    /complement(1..392)
    /gene="recA"
    /complement(1..392)
    /gene="recA"
    /note="MLCB33.01c recA, (partial) contains PS00017
    ATP/GTP-binding site motif A"
    /codon_start=1
    /transl_table=11
    /product="RecA"
    /protein_id="CAB08125.1"
    /db_xref="PID:e313311"
    /db_xref="PID:g2062105"
    /db_xref="GI:2062105"
    /db_xref="SPTREMBL:O05556"
    /translation="NAQVPDREKALEMAQTEKNYKGVNRLGDEMCOPTSVIPTG
    STALDVALGGLPRGRIVEIYGPSSGKTTVALHAVANAQVGVGAFFIDAEHALEP
    EYAKLGVDTDSLLVSQPDGQALETGS"
  5..9
    /gene="recA"
    /note="cagtc missing in entry EM_BA:ML019 U00019"
  8..6037
    /note="overlap with positions, 3000..36033 EM_BA:ML019
    U00019 Mycobacterium leprae cosmid B2235, note conflict
    with this sequence at the 5' end"
    /complement(171..194)
    /gene="recA"
    /note="PS00017 ATP/GTP-binding site motif A"
    /complement(400..404)
    /note="aggag, potential rbs upstream of recA"
    /complement(760..1938)
    /gene="MLCB33.02c"
    /complement(760..1938)
    /gene="MLCB33.02c"
    /note="MLCB33.02c, possible glycosyl transferase, len: 392
    aa, identical to Q49841 U2235C except more upstream
    initiation codon chosen on the basis of similarity eg to
    CRTX_ERWHE Q01330 zeaxanthin glucosyl transferase (413
    aa), fasta scores, opt: 233, E(): 1.4e-07, (26.1% identity
    in 410 aa overlap)"
    /codon_start=1
    /transl_table=11
    /product="unknown"
    /protein_id="CAB08126.1"
    /db_xref="PID:e313312"
    /db_xref="PID:g2052098"
    /db_xref="GI:2052098"
    /db_xref="SPTREMBL:Q49841"
    /translation="MVHYMRVAVAGPDGPHGFPAIALCQRFAGAGDPTTLFTGVECV
    DIARAAGVETALLDGLVAVDDVDAGARHRRAAQMAVNVFVLRDLADPLIVSDVIT
    AAGGMAELLGTPWELNPHPLPLPSKGLPIPGSLAGVGIHGRMDRTWRLTARS
    WRAGLRERAARVAVQIGLPACDSGGLRLRIATLPALVEVRPDWPAEAVVVGPDHFEPTD

```

gene

CDS

```

RVLIQIPSGSPVIVVAPSTALTIGSMAQTALDVLPGDTLPSSGSRVVVSRLLNGPDLT
VPSMAVAGLVHOTIKLLAHADVVICGGHGMVAKTLLAGVPLVVPVPGGQOWENARVV
ROGSQALRPLTAALLAAVNKVLSSPDYREAAQRAAISIAGVADPVQCHEALGLAG
1888..2424
  /gene="MLCB33.03"
  /note="MLCB33.03, unknown, len: 178 aa, identical to
  Q49850 B2235_F3_140"
  /codon_start=1
  /transl_table=11
  /product="unknown"
  /protein_id="CAB08127.1"
  /db_xref="PID:e313313"
  /db_xref="PID:g2052099"
  /db_xref="GI:2052099"
  /db_xref="SPTREMBL:Q49850"
  /translation="MPGTWPGDHSNAHVYVHAIPLCHVSPKLVAIAELTEASVQVGE
  NIRTVEVFLAALQDAGLRNRIRDVGROPRVYQNVGLPTIHGRSKTITLWRKMADICIGF
  EIKIHRIAVAIAVICERADAVIVGPLMWFVCGTTEYQNKRMILWRNYFDLFLFK
  ATMRSLVALRIPSLNAAF"
  /complement(1946..1949)
  /gene="MLCB33.03"
  /note="ggtg, potential rbs upstream of MLCB33.02c"
  /complement(3317..3688)
  /gene="MLCB33.04c"
  /complement(3317..3688)
  /gene="MLCB33.04c"
  /note="MLCB33.04c, unknown, len: 123 aa, identical to
  Q49833 B2235_C1_148"
  /codon_start=1
  /transl_table=11
  /product="unknown"
  /protein_id="CAB08128.1"
  /db_xref="PID:e313314"
  /db_xref="PID:g2052100"
  /db_xref="GI:2052100"
  /db_xref="SPTREMBL:Q49833"
  /translation="MNVSGQWRALMRRLDITDLSLLAKKISVLDPRALLRR
  RRRALRWGLIFGCGLFWGVVTTALLAANGWFFLLLIQITGSIADVLAQIPATLLLRVW
  LRSEPLTWRPNVRRLLPPG"
  /complement(3692..3697)
  /note="gaggtg, potential
  /complement(3726..4070)
  /gene="MLCB33.05c"
  /complement(3726..4070)
  /gene="MLCB33.05c"
  /note="MLCB33.05c, unknown, len: 114 aa, identical to
  Q49845 B2235_C3_214, similar to 35KD_MYCTU P31511 35 kd
  protein, M. tuberculosis (270 aa), fasta scores, opt: 582,
  E(): 8.5e-30, (66.2% identity in 145 aa overlap)
  MLCB33.05c and MLCB33.06c appear frameshifted with respect
  to 35KD_MYCTU"
  /codon_start=1
  /transl_table=11
  /product="unknown"
  /protein_id="CAB08116.1"
  /db_xref="PID:e313315"
  /db_xref="PID:g2052101"
  /db_xref="GI:2052101"
  /db_xref="SPTREMBL:Q49845"
  /translation="MQEQVSASLRSMSEIAPVPGDPTLDDEVREKTIERRYVNALGAEL
  AQGSVOGRMIETEQAGVQMGHLRLEQIRASMLSOALSITGGTAAPFGGTPIASINQA
  SDVHVVEKSLGO"
  /complement(4054..4557)
  /gene="MLCB33.06c"
  /complement(4054..4557)
  /gene="MLCB33.06c"
  /note="MLCB33.06c, unknown, possible coiled coil protein,
  len: 167 aa, identical to Q49840 B2235_C2_187 and similar
  to 35KD_MYCTU P31511 35 kd protein, M. tuberculosis (270
  aa), fasta scores, opt: 789, E(): 0, (85.0% identity in

```

gene

CDS


```

147 aa overlap);MLCB33.05c and MLCB33.06c appear
frameshifted with respect to 35KD_MYCTU*
/codon_start=1
/transl_table=11
/product="unknown"

Query Match      82.0%; Score 16.4; DB 1; Length 42224;
Best Local Similarity 94.4%; Pred. NO. 2e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      2 GGCATTGATCTGGTTTCAT 19
        |||||
Db 12621 GGCATTGATCTGGATCAT 12604

```

Search completed: September 18, 1999, 15:48:49
Job time: 1409 sec

Result No.	Query	Score	Query			DB	ID	Description
			Match	Length	%			
1	20	100.0	20	1	T13885	Primer for detecti		
2	20	100.0	20	1	V32391	Interleukin-1-beta		
3	20	100.0	20	1	V60232	PCR primer used to		
4	20	100.0	20	1	X16610	Interleukin 1 (441		
5	20	100.0	14690	1	X22303	Human IL-1ra BAC c		
6	20	100.0	10620	1	X02996	Human IL-1ra BAC c		
7	16.8	84.0	2854	1	V32591	Schwannomin-bindin		
8	16.8	84.0	558	1	V86039	EST clone B238. Ne		
9	16.4	82.0	30246	1	V74367	Staphylococcus aur		
10	16	80.0	6914	1	Q55757	Enterobacter cloaci		
11	15.8	79.0	2220	1	V21461	Rhodospiridium tor		
12	15.8	79.0	1738	1	V21460	Rhodospiridium tor		
13	15.8	79.0	930	1	V31298	Rhodospiridium tor		
14	15.4	77.0	3004	1	T12337	E. coli J96 pathoge		
15	15.4	77.0	714	1	X34151	A. cellulolyticus		
16	15.4	77.0	714	1	X34127	Mycobacterium spec		
17	15.4	77.0	750	1	X34128	Mycobacterium spec		
18	15.2	76.0	1073	1	N82405	Mycobacterium spec		
19	15.2	76.0	657	1	N82406	pgTR200-DNA insert		
20	15.2	76.0	1223	1	N82407	pgTR200-DNA insert		
21	15.2	76.0	657	1	Q30204	Glyb clone contg.		
22	15.2	76.0	1222	1	Q30205	Rat liver glutathi		
23	15.2	76.0	1990	1	Q44294	Rat brain glutathi		
24	15.2	76.0	1684	1	Q34673	pectin esterase cl		
25	15.2	76.0	1440	1	Q94172	Rinder-pest virus		
26	15.2	76.0	2943	1	T16480	Neuropeptide Y/pep		
27	15.2	76.0	110000	1	T58840_4	SAB virus GB glyco		
28	15.2	76.0	490	1	V13078	Continuation (5 of		
29	15.2	76.0	2097	1	V29755	Cellulase activity		
30	15.2	76.0	1678	1	V40278	Zea mays soluble s		
31	14.8	74.0	16079	1	N70753	Rat equilibrative		
32	14.8	74.0	885	1	Q63682	Sequence of segmen		
33	14.8	74.0	1316	1	Q56831	Partial cDNA encod		
34	14.8	74.0	3077	1	V79920	Full-length human		
35	14.8	74.0	1908	1	T63269	E. coli accB and a		
36	14.8	74.0	1454	1	V23913	Gene for firefly l		
37	14.8	74.0	1474	1	V32914	Plant CAD enzyme D		
38	14.8	74.0	1193	1	V32646	Plant CAD enzyme D		
39	14.8	74.0	885	1	X82583	DNA encoding Bce44		
40	14.8	74.0	885	1	X01871	Thyroid hormone re		
41	14.8	74.0	3076	1	X02987	Human TR-interacti		
42	14.8	74.0	16075	1	V99811	Human IL-1ra BAC c		
43	14.4	72.0	1161	1	N70828	Gum gene cluster. DNA encoding human		

PI Duff G, Rennie . Richardson

WLT, 20 240635/21.

PT Predicting increased risk of sight-threatening diabetic retinopathy
PT - comprises identifying genetic polymorphism pattern for genes
PT IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
PT symptoms occur
PS Claim 2; Page 33; 41pp; English.
CC Interleukin-1-beta (IL-1-beta) primers 1 and 2 (V32392) were used to
CC amplify the IL-1-beta gene region to identify single base variation
CC polymorphism of C/T at base 511. The invention claims to provide a
CC method for predicting the risk of sight threatening diabetic retinopathy.
CC The method involves isolating DNA from a patient and determining the DNA
CC polymorphism pattern of the genes that code for interleukin-1-alpha,
CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
CC identified is then compared with controls of known DNA polymorphism
CC patterns thereby identifying patients carrying a genetic polymorphism
CC associated with increased risk of sight threatening diabetic retinopathy.
CC The method may be able to identify diabetic patients at risk before the
CC clinically detectable disorders occur. Polymorphism pattern
CC determination of IL genes involved PCR reactions using primers V32389-
CC V32398. The method is also claimed to be useful in conjunction with
CC identification of other genes associated with sight threatening diabetic
CC retinopathy in genomic DNA and therefore, in identifying diabetic
CC patients expressing multiple risk patterns.
CC Sequence 20 BP; 3 A; 4 C; 5 G; 8 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.24; 0; Indels 0; Gaps 0;
Matches 20; Conservative 0; Mismatches 0;

QY 1 TGGCATTGATCTGGTTCATC 20

DB 1 TGGCATTGATCTGGTTCATC 20

RESULT 3

V60232 ID V60232 standard; DNA; 20 BP.

AC V60232;

DT 25-NOV-1998 (first entry)

DE PCR primer used to amplify interleukin-1B (IL-1B).

KW Interleukin-1B; IL-1B; predisposition; coronary artery disease;

KW screen; PCR primer; ss.

OS Synthetic.

PN W09840517-A1.

PD 17-SEP-1998.

PF 09-MAR-1998; U04725.

PR 10-MAR-1997; US-813456.

PA (MEDI-) MEDICAL SCI SYSTEMS INC.

PI Crossman DC, Duff GW, Francis SE;

WPI: 98-520829/44.

DT Detection of predisposition to coronary artery disease - by
PT comparative measurement of levels of expression of alleles from the
PT interleukin 1 locus

PS Claim 6; Page 15; 22pp; English.

CC PCR primers V60232-33 were used to amplify alleles associated with
CC interleukin-1B (IL-1B). The specification describes a method for
CC determination of a patient's predisposition to coronary artery
CC disease. The method comprises comparing an allele with a second
CC allele which is predictive of coronary artery disease, where
CC similarity between the first and second alleles indicates a
CC predisposition to coronary artery disease. The method is
CC used to genotype an individual's interleukin (IL)-1 loci, the
CC overexpression of which correlates with coronary artery disease. The
CC method is used to screen a patients' predisposition to coronary
CC artery disease.
CC Sequence 20 BP; 3 A; 4 C; 5 G; 8 T;

Query Match 100.0%; Score 20; DB 1; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20

DB 1 TGGCATTGATCTGGTTCATC 20

RESULT 4

X16615

ID X16615 standard; DNA; 20 BP.

AC X16615;

DT 29-APR-1999 (first entry)

DE Interleukin 1 (44112332) haplotype PCR primer #9.

KW Interleukin 1; IL-1; haplotype; inflammatory disorder; alopecia areata;

KW coronary artery disease; osteoporosis; nephropathy; diabetes mellitus;

KW Graves disease; systemic lupus erythematosus; lichen sclerosis;

KW ulcerative colitis; PCR primer; ss.

OS Synthetic.

OS Homo sapiens.

PN W09854359-A1.

PD 03-DEC-1998.

PF 21-MAY-1998; G01481.

PR 29-MAY-1997; GB-011040.

PA (CAMP/) CAMP N J.

PA (COXA/) COX A.

PA (DGIO/) DE GIOVINE F S.

PA (DUFF/) DUFF G.

PI Camp NJ, Cox A, De Giovine FS, Duff G;

WPI: 99-080814/07.

PT New method of determining a patient's susceptibility to inflammatory

PT disorders - by detecting the presence of an IL-1 (44112332)

PT haplotype, useful in designing treatment strategies that modulate

PT the activity of proteins produced by the IL-1 gene cluster

PS Claim 3; Page 33; 49pp; English.

CC A method has been developed for determining a patient's susceptibility

CC to an inflammatory disorder. The method comprises the detection of an

CC interleukin 1 (IL-1) (44112332) haplotype in a sample obtained from the

CC patient, where its presence indicates susceptibility to an inflammatory

CC disorder. X16607 to X16631 represent PCR primer used in the method for

CC detecting the IL-1 (44112332) haplotype. The method provides kits for

CC the early prediction of a patient's susceptibility to inflammatory

CC disorders, including coronary artery disease, osteoporosis, nephropathy

CC in diabetes mellitus, alopecia areata, Graves disease, systemic lupus

CC erythematosus, lichen sclerosis and ulcerative colitis. The detection of

CC alleles of the haplotype can be applied to particular inflammatory

CC disorders, comprising diabetic retinopathy, juvenile chronic arthritis,

CC psoriasis, and insulin dependent diabetes. The identification of a

CC disease-associated haplotype enables determination of which alleles are

CC causative, and this information is useful in designing treatment

CC strategies, including gene therapy and treatment using various agents

CC that modulate the activity of proteins produced by the IL-1 gene cluster.

CC Some alleles from the IL-1 gene cluster are associated with particular

CC inflammatory diseases, and insufficient IL-1 production appears to act

CC centrally in the pathology of these diseases. Therefore, the use of IL-1

CC gene clusters is useful in determining genetic susceptibility to

CC inflammatory diseases, including those with a multifactorial etiology

CC with a polygenic component.
CC Sequence 20 BP; 3 A; 4 C; 5 G; 8 T;

Query Match 100.0%; Score 20; DB 1; Length 20;

Best Local Similarity 100.0%; Pred. No. 0.24;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20

DB 1 TGGCATTGATCTGGTTCATC 20

RESULT 5

X22303/c

ID X22303 standard; DNA; 14690 BP.

AC X22303;

DT 22-JUN-1999 (first entry)

DE Human IL-1ra BAC contiguous DNA sequence 96.

KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;

KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
PN WO9906426-A1.
PD 11-FEB-1999. U16102.
PF 03-AUG-1998; US-091650.
PR 02-JUL-1998; US-091650.
PR 04-AUG-1997; US-054646.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Pan Y.
DR WPI: 99-153692/13.

PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
PS Example 5; Figure 4; 226pp; English.
CC X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC containing alternatively spliced forms of human IL-1ra. Such fragments
CC are used in the method of the invention which describes the isolation of
CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC member of the cytokine superfamily that is expected to inhibit
CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC also bind to a new receptor so could regulate other cellular processes
CC associated with acute or chronic inflammation, e.g. asthma, chronic
CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent
CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 14690 BP; 3810 A; 3245 C; 3299 G; 4327 T;

Query Match 100.0%; Score 20; DB 1; Length 14690;
Best Local Similarity 100.0%; Pred. No. 0.59;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGTTTCATC 20
|||||
DB 10520 TGGCATTGATCTGTTTCATC 10501

RESULT 6

ID X02996 standard; DNA; 10620 BP.
AC X02996;
DT 22-JUN-1999 (first entry)
DE Human IL-1ra BAC contiguous DNA sequence 41.
KW Tango-77; human; cytokine superfamily; inflammation; inhibition;
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
PN WO9906426-A1.
PD 11-FEB-1999.
PF 03-AUG-1998; U16102.
PR 02-JUL-1998; US-091650.
PR 04-AUG-1997; US-054646.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Pan Y.

DR WPI: 99-153692/13.
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
PS Example 5; Figure 3; 226pp; English.
CC X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC containing alternatively spliced forms of human IL-1ra. Such fragments
CC are used in the method of the invention which describes the isolation of
CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC member of the cytokine superfamily that is expected to inhibit
CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC also bind to a new receptor so could regulate other cellular processes
CC associated with acute or chronic inflammation, e.g. asthma, chronic
CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent

CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 10620 BP; 2746 A; 2407 C; 2411 G; 3049 T;

Query Match 100.0%; Score 20; DB 1; Length 10620;
Best Local Similarity 100.0%; Pred. No. 0.56;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGTTTCATC 20
|||||
DB 6360 TGGCATTGATCTGTTTCATC 6341

RESULT 7

V32591/c
ID V32591 standard; cDNA; 2854 BP.
AC V32591;
DT 28-SEP-1998 (first entry)
DE Schwannomin-binding protein cDNA.
KW Schwannomin-binding protein; SBP; human; neurofibromatosis 2;
KW NF2; tumour suppressor protein; brain tumour; glioma; schwannoma;
KW meningioma; sarcoma; melanoma; ependymoma; cataract; hamartoma;
KW therapy; diagnosis; transgenic animal; signal transduction;
KW cell proliferation; translation; ss; ds.
OS Homo sapiens.

FF Key Location/Qualifiers
FT CDS 45..2786 /*tag= a
FT WO9821333-A2.
PN 22-MAY-1998.
PF 14-NOV-1997; U21005.
PF 15-NOV-1996; US-030987.
PR (CEDA-) CEDARS SINAI MEDICAL CENT.
PI Pulst SM, Scoles DR;
DR WPI: 98-297937/26.
DR P-PSDB; W48928.
PT New nucleic acid encoding schwannomin-binding protein - useful for,
PT e.g. diagnosis and treatment of brain tumours and in drug screening
PS Claim 3; Page 53-57; 75pp; English.

CC This cDNA sequence encodes a human schwannomin-binding protein
CC (SBP) (see W48928) that has been found to have functional
CC involvement in the initiation of translation and is suggested to
CC serve a role in the final steps of a signal transduction cascade
CC affecting cell division and proliferation. The cDNA was isolated
CC from a human adult brain cDNA library using a yeast two-hybrid
CC method with schwannomin isoform 1 as bait. Novel SBPs of the
CC invention (see W48927-31) bind to the neurofibromatosis 2 (NF2)
CC tumour suppressor protein schwannomin, and have a role in cellular
CC pathways that effect cell division and proliferation. They provide
CC targets for treating a broad variety of cancer pathologies, such as
CC glioma, carcinoma, sarcoma, melanoma and hamartoma. SBPs are used
CC in a claimed method for modulating the activity of an oncogenic
CC protein. SBP nucleic acids, oligonucleotides (including antisense),
CC vectors containing these, transfected host cells and SBP polypeptides
CC and antibodies of the invention can be used to screen for potential
CC antagonists and agonists of SBP. These can be used to treat brain
CC tumours, such as glioma, schwannoma, meningioma and ependymoma, as
CC well as cataracts and retinal disorders. Oligonucleotides that
CC specifically hybridise with nucleic acids of the invention (see
CC V32590-94) are used to identify or quantify nucleic acids encoding
CC SBP, e.g. in library screening, and may also be used to detect SBP
CC genes mutations, particularly for diagnosis of cancer. Antisense
CC nucleic acids are used to inhibit SBP expression. Single-stranded
CC primers based on SBP nucleic acids are also claimed. Transgenic
CC animals expressing exogenous nucleic acid encoding a SBP are models
CC for studying activity of SBP and for drug screening.
SQ Sequence 2854 BP; 767 A; 756 C; 807 G; 524 T;

Query Match 84.0%; Score 16.8; DB 1; Length 2854;
Best Local Similarity 90.0%; Pred. No. 17;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;


```

FT      /*tag= n
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence."
FT      misc_feature 26941..27000
FT      /*tag= o
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence."
FT      misc_feature 28741..28800
FT      /*tag= p
FT      /note= "these bases represent a line of missing text in
FT      the sequence listing in the specification. They
FT      are included to maintain the nucleotide numbering
FT      given in the specification for this DNA sequence."
PN      EP-786519-A2.
PD      30-JUL-1997.
PF      07-JAN-1997; 100117.
PR      05-JAN-1996; US-009861.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI      Rosen CA;
DR      WPI; 97-374922/35.
PT      Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT      stored on computer readable medium and used in the production of
PT      anti-S.aureus vaccines
PS      Claim 1; Page 422-438; 3271pp; English.
CC      This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC      of the invention. The DNA sequences are recorded on a computer readable
CC      medium, preferably selected from a floppy or hard disk, random access
CC      memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC      the S.aureus DNA sequences allows putative functions to be assigned so
CC      that protein-encoding or regulatory regions of commercial, therapeutic or
CC      industrial importance can be obtained. Specifically, sequences which are
CC      likely to encode antigens have been identified and these polypeptides can
CC      be used in a vaccine composition against S.aureus infection. The
CC      polypeptides can also be used in a kit for the immunodetection of
CC      S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC      including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC      skin and surgical wound infections, scalded skin syndrome, toxic shock
CC      syndrome, etc. Organisms transformed with the DNA sequences can be used
CC      for recombinant production of the polypeptides. The new DNA sequences
CC      (and their fragments) are useful as primers or probes for isolating
CC      homologues of any of the S.aureus DNA sequences contained on the
CC      computer readable medium.
SQ      Sequence 30246 BP; 9663 A; 4436 C; 6386 G; 8791 T;

Query Match 82.0%; Score 16.4; DB 1; Length 30246;
Best Local Similarity 94.4%; Pred. No. 37;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TGCATTGATCTGGTTCA 18
Db      1324 TGCATTGATCTGGTTCA 1307
||||||| |||||

RESULT 10
Q55757/c
ID      Q55757 standard; DNA; 6914 BP.
AC      Q55757;
DT      26-SEP-1995 (first entry)
DE      Enterobacter cloacae genomic probe ET-49.
KW      Probe: S.aureus; S.epidermis; E.faecalis; P.aeruginosa; E.coli;
KW      K.pneumoniae; E.cloacae; clinical sample; ds.
OS      Enterobacter cloacae.
PN      WO9401583-A.
PD      20-JAN-1994.
PF      07-JUL-1993; J00936.
PR      07-JUL-1992; JP-179719.
PA      (FUSO ) FUSO PHARM IND LTD.

```

```

PA      (OHNO/) OHNO T.
PI      Eda S, Matsuhisa A, Ohno T, Uehara H;
DR      WPI; 94-035086/04.
PT      Probe for identifying bacteria causing infectious disease -
PT      consists of a DNA fragment obtained by HindIII cleavage of the
PT      pathogenic bacterial genomic DNA
PS      Claim 7; Page 78-82; 100pp; Japanese.
CC      The nucleotide sequence of a 6914 bp probe obtained by digestion of
CC      Enterobacter cloacae genomic DNA with the restriction enzyme HindIII.
CC      The probes (Q55133-48 and Q55752-58) represent other probes derived,
CC      by HindIII digestion of the genomes of Staphylococcus aureus,
CC      S.epidermis, Enterococcus faecalis, Pseudomonas aeruginosa, E.coli,
CC      Klebsiella pneumoniae or Enterobacter cloacae. The probes can be used
CC      to detect their respective microorganisms in clinical samples.
SQ      Sequence 6914 BP; 1420 A; 1937 C; 1807 G; 1750 T;

Query Match 80.0%; Score 16; DB 1; Length 6914;
Best Local Similarity 100.0%; Pred. No. 48;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GGCATTGATCTGGTTC 17
Db      901 GGCATTGATCTGGTTC 886
||||||| |||||

RESULT 11
V21461/c
ID      V21461 standard; cDNA; 2220 BP.
AC      V21461;
DT      17-AUG-1998 (first entry)
DE      Rhodospiridium toruloides cephalosporin esterase genomic DNA.
DE      Cephalosporin esterase; antibiotic; ds.
OS      Rhodospiridium toruloides strain ATCC 10657.
FT      Key
FT      Location/Qualifiers
FT      CDS
FT      77..2108
FT      /*tag= a
FT      /note= "contains 5 introns"
FT      exon
FT      1..694
FT      /*tag= b
FT      /number= 1
FT      695..757
FT      /*tag= c
FT      /number= 1
FT      758..877
FT      /*tag= d
FT      /number= 2
FT      878..950
FT      /*tag= e
FT      /number= 2
FT      951..1283
FT      /*tag= f
FT      /number= 3
FT      1284..1342
FT      /*tag= g
FT      /number= 3
FT      1343..1376
FT      /*tag= h
FT      /number= 4
FT      1377..1439
FT      /*tag= i
FT      /number= 4
FT      1440..1652
FT      /*tag= j
FT      /number= 5
FT      1652..1708
FT      /*tag= k
FT      /number= 5
FT      1709..2220
FT      /*tag= l
FT      WO9812345-A1.
PN      26-MAR-1998.
PD      11-SEP-1997; U16193.

```

PR 18-SEP-1996; US-026929.
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 PI Burnett WV, Politino M, Romancik G, Tonzi SM, Usher JJ;
 DR WPI: 98-230321/20.
 DR P-PSDB: W54159.
 PT Isolated cephalosporin esterase gene - obtained from Rhodospiridium
 PT toruloides, used to obtain products which can be used for hydrolysis
 PT of cephalosporins
 PS Claim 8; Fig 6A-B; 38pp; English.
 CC This nucleic acid molecule comprises genomic DNA of Rhodospiridium
 CC toruloides ATCC 10657 coding for cephalosporin esterase (see
 CC W54159). It was isolated by using oligonucleotide probes (see
 CC T99709-12) based on the N-terminal sequence of the mature enzyme to
 CC screen Southern blots of R. toruloides chromosomal DNA. A cDNA
 CC clone (see V21460) is also provided. These isolated nucleic acid
 CC molecules can be used for the recombinant production of
 CC cephalosporin esterase using claimed expression vectors and claimed
 CC host cells (preferably Escherichia coli, R. toruloides,
 CC Cephalosporium acremonium or Penicillium chrysogenum). The
 CC cephalosporin esterase can be used for the hydrolysis of the 3'
 CC acetyl groups of cephalosporins. The polypeptide (even if
 CC catalytically inactive) can also be used for the production of
 CC antibodies for use in detection and purification.
 CC Sequence 220 BP; 389 A; 814 C; 554 G; 463 T;
 SQ

Query Match 79.0%; Score 15.8; DB 1; Length 2220;
 Best Local Similarity 89.5%; Pred. No. 52;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
 | ||| ||||| |||||
 Db 875 TCGCAATGATCTGGTTCAT 857

RESULT 12
 V21460/c
 ID V21460 standard; cDNA; 1738 BP.
 AC V21460;
 DT 17-AUG-1998 (first entry)
 DE Rhodospiridium toruloides cephalosporin esterase cDNA.
 KW Cephalosporin esterase; antibiotic; ss.
 OS Rhodospiridium toruloides strain ATCC 10657.
 FH Key Location/Qualifiers
 FT CDS 11..1729
 FT /*tag= a
 FT sig_peptide 11..94
 FT /*tag= b
 FT mat_peptide 95..1726
 FT /*tag= c
 PN W0981245-A1.
 PD 26-MAR-1998.
 PF 11-SEP-1997; U16193.
 PR 18-SEP-1996; US-026929.
 PA (BRIM) BRISTOL-MYERS SQUIBB CO.
 PI Burnett WV, Politino M, Romancik G, Tonzi SM, Usher JJ;
 DR WPI: 98-230321/20.
 DR P-PSDB: W54159.
 PT Isolated cephalosporin esterase gene - obtained from Rhodospiridium
 PT toruloides, used to obtain products which can be used for hydrolysis
 PT of cephalosporins
 PS Claim 7; Fig 5; 38pp; English.
 CC This cDNA clone codes for the cephalosporin esterase (see W54159)
 CC of Rhodospiridium toruloides ATCC 10657. It was isolated from
 CC total RNA of R. toruloides by first-strand cDNA preparation and
 CC amplification using a gene-specific primer and an adapter primer.
 CC A genomic DNA sequence (see V21461), which contains 5 introns, is
 CC also provided. These isolated nucleic acid molecules can be used
 CC for the recombinant production of cephalosporin esterase using
 CC claimed expression vectors and claimed host cells (preferably
 CC Escherichia coli, R. toruloides, Cephalosporium acremonium or
 CC Penicillium chrysogenum). The cephalosporin esterase can be used
 CC for the hydrolysis of the 3' acetyl groups of cephalosporins. The

CC polypeptide (even if catalytically inactive) can also be used for
 CC the production of antibodies for use in detection and purification.
 SQ Sequence 1738 BP; 328 A; 639 C; 443 G; 328 T;
 0;

Query Match 79.0%; Score 15.8; DB 1; Length 1738;
 Best Local Similarity 89.5%; Pred. No. 50;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
 | ||| ||||| |||||
 Db 746 TCGCAATGATCTGGTTCAT 728

RESULT 13
 V31298/c
 ID V31298 standard; DNA; 930 BP.
 AC V31298;
 DT 01-OCT-1998 (first entry)
 DE E. coli J96 pathogenicity island contig #112.
 KW PAI; pathogenicity island; uropathogenic E. coli detection; PAI IV; pheR;
 KW PAI V; pheV; vaccine; protective immune response; ds.
 OS Escherichia coli.
 PN W09822575-A2.
 PD 28-MAY-1998.
 PF 21-NOV-1997; U21347.
 PR 14-OCT-1997; US-061953.
 PR 22-NOV-1996; US-031626.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (UYWI-) UNIV WISCONSIN.
 PI Choi GH, Dillon PJ, Welch RA;
 DT WPI: 98-312461/27.
 PT New isolated uropathogenic E. coli nucleotide sequences - used to
 PT develop products for the detection of pathogenic E. coli and to
 PT elicit an immune response to pathogenic E. coli
 PS Claim 21; Page 224; 250pp; English.
 CC This sequence represents a E. coli strain J96 contig containing
 CC pathogenicity island (PAI) sequences, and represents a nucleic acid
 CC molecule of the invention. PAIs are large fragments of DNA which comprise
 CC pathogenicity determinants. The sequences of the invention taken from
 CC PAI IV and PAI V. PAI IV is located at approximately 64 min (near pheV)
 CC on the E. coli chromosome and is greater than 170 kb. PAI V is located at
 CC approximately 94 min (at pheR) on the E. coli chromosome and is
 CC approximately 160 kb in size. Antibodies specific to the proteins encoded
 CC by the PAI open reading frames of the invention can be used in kits to
 CC detect uropathogenic E. coli. The proteins are used in vaccines to elicit
 CC a protective immune response in an animal to the uropathogenic E. coli
 CC strain J96.
 SQ Sequence 930 BP; 319 A; 196 C; 162 G; 249 T;
 0;

Query Match 79.0%; Score 15.8; DB 1; Length 930;
 Best Local Similarity 89.5%; Pred. No. 46;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
 | ||||| |||||
 Db 213 TAGCATTGATCTGGTTCAT 195

RESULT 14
 T12337/c
 ID T12337 standard; DNA; 3004 BP.
 AC T12337;
 DT 08-OCT-1996 (first entry)
 DE A. cellulolyticus E1 endoglucanase gene.
 KW E1 endoglucanase; cellulase; cellulose; saccharification; ethanol;
 KW ss.
 OS Acidothermus cellulolyticus.
 FH Key Location/Qualifiers
 FT misc_difference 200
 FT /*tag= a
 FT /note= "base g at position 200 is shown in Fig 1,
 FT

Mon Sep 20 14:16:48 1999

FT but is missing in the sequence given on
 FT page 22 of the specification"
 FT 710. .725
 FT /tag= b
 FT /note= "putative upstream regulatory sequence"
 FT 772. .779
 FT /tag= c
 FT 781. .790
 FT /tag= d
 FT /rpt_type= DIRECT
 FT 795. .804
 FT /tag= e
 FT /rpt_type= DIRECT
 FT 810. .817
 FT /tag= f
 FT /rpt_type= DIRECT
 FT 824. .2512
 FT /tag= g
 FT signal_peptide 824. .946
 FT /tag= h
 FT signal_peptide 863. .946
 FT /tag= i
 FT /note= "alternative signal sequence"
 FT 947. .2509
 FT /tag= j
 FT terminator 2514. .2560
 FT /tag= k

PN WO9602551-A1.
 PD 01-FEB-1996.
 PF 14-JUL-1995; U08868.
 PR 26-SEP-1989; US-412434.
 PR 27-JAN-1992; US-826089.
 PR 21-SEP-1993; US-125115.
 PR 15-JUL-1994; US-276213.
 PR (MIDE) MIDWEST RES INST.
 PA Hammel ME, Laymon RA, Thomas SR;
 DR WPI; 96-105843/11.
 DR P-PSDB; R89927.
 DR New isolated DNA encoding endoglucanase - obtd from Acidothermus
 PT cellulolyticus, used for prodn of the enzyme for use in cellulose
 PT hydrolysis.
 PS Claim 4; Page 22-24; 34pp; English.
 CC The EI endoglucanase gene (Ti2337) of Acidothermus cellulolyticus
 CC codes for an enzyme, the mature form of which (R89927) can hydrolyse
 CC cellulose and xylan. The gene was obtd. by ligating Sau3A fragments
 CC of A. cellulolyticus genomic DNA into EMBL3 lambda phage arms,
 CC packaging the phage DNA, plating on E.coli LE392 agar contg. CMC,
 CC staining with Congo Red, and isolating DNA from positive clones.
 CC Alternative signal sequences were identified (see also R89928-29).
 CC The gene can be incorporated into a vector and used for large-scale,
 CC low-cost prodn. of recombinant EI endoglucanase, pref. using
 CC Saccharomyces, Zymomonas or E.coli hosts. The enzyme is useful
 CC for the saccharification of cellulosic biomass for fermentation to
 CC ethanol.
 SQ Sequence 3004 BP; 604 A; 931 C; 938 G; 531 T;

Query Match 77.0%; Score 15.4; DB 1; Length 3004;
 Best Local Similarity 94.1%; Pred. No. 84;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTC 17
 |||||
 Db 2824 TGGCATTGATCTGGTTC 2808

RESULT 15
 X34151/C
 ID X34151 standard; DNA: 714 BP.
 AC X34151;
 DT 06-JUL-1999 (first entry)
 DE Mycobacterium species nucleic acid sequence 29.
 KW Secreted protein; Mycobacterium; primer; PCR; amplification; probe;

KW hybridisation; detection; vaccine; immunisation; infection; ss.
 OS Mycobacterium sp.
 PN WO9909186-A2.
 PD 25-FEB-1999.
 PF 14-AUG-1998; F01813.
 PR 11-SEP-1997; FR-011325.
 PR 14-AUG-1997; FR-010404.
 PA (INSP) INST PASTEUR.
 PI Gicquel B, Lim EM, Pellicic V, Portnoi D, Goguet de la Salmoniere Y,
 PI Guigueno A;
 DR WPI; 99-181045/15.
 PT Mycobacterial DNA vectors containing reporter constructs - for
 PT identifying coding or promoter sequences involved in
 PT infection-associated protein expression
 PS Claim 22; Fig 29; 309pp; French.
 CC Sequences X34001-X34252 represent nucleic acids encoding secreted
 CC proteins from various Mycobacterium species microorganisms. The
 CC nucleotide sequences can be used as primers and probes for methods
 CC for detecting and identifying mycobacteria, especially belonging to
 CC the M. tuberculosis complex. The encoded proteins can be used in
 CC vaccines for immunisation against a bacterial or viral infection.
 SQ Sequence 714 BP; 153 A; 212 C; 225 G; 124 T;

Query Match 77.0%; Score 15.4; DB 1; Length 714;
 Best Local Similarity 94.1%; Pred. No. 89;
 Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 GGCATTGATCTGGTTC 18
 |||||
 Db 180 GGCATTGATCTGGTTC 164

Search completed: September 18, 1999, 05:27:18
 Job time: 1683 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:32 ; Search time 1405 Seconds
(without alignments)
28.079 Million cell updates/sec

Title: US-09-037-472-3

Perfect score: 20
Sequence: 1 TGGCATTGATCTGGTTCATC 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database : EST:*

- 1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *
45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *

- 54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	17.4	87.0	359	21	F07416	F07416 HSC25B041 n
2	17.4	87.0	429	30	AA214464	AA214464 z989b01.r
3	17.4	87.0	430	45	AI365968	AI365968 a091c06.x
4	16.8	84.0	516	20	T27501	T27501 EST27116 WA
5	16.8	84.0	296	20	T34213	T34213 EST64297 Hu
6	16.8	84.0	494	30	AA265846	AA265846 mz53d03.r
7	16.8	84.0	443	31	AA305985	AA305985 EST176975
8	16.8	84.0	311	35	AA581691	AA581691 nn52a04.s
9	16.8	84.0	384	48	AI600819	AI600819 486072C05
10	16.8	84.0	500	49	AL045670	AL045670 DKF2p434A
11	16.8	84.0	421	49	AL043103	AL043103 DKF2p434C
12	16.8	84.0	451	51	AI722423	AI722423 fc27a12.x
13	16.8	84.0	867	53	HSM007819	AI042969 Homo sapi
14	16.8	84.0	708	54	HSM009381	AI04531 Homo sapi
15	16.8	84.0	539	54	HSM010310	AL045460 Homo sapi
16	16.8	84.0	540	54	HSM010808	AI045938 Homo sapi
17	16.4	82.0	360	20	D39361	D39361 RICS0022A R
18	16.16	80.0	620	50	AU060199	AU060199 AU060199
19	15.8	79.0	360	20	D37412	D37412 CELK0045AX
20	15.8	79.0	447	23	R89462	R89462 yq01b10.r1
21	15.8	79.0	325	24	D65351	D65351 CELK055B9R
22	15.8	79.0	474	26	W95680	W95680 ze07h05.s1
23	15.8	79.0	330	27	AA023517	AA023517 mh77c09.r
24	15.8	79.0	394	34	AA492755	AA492755 vi76g02.r
25	15.8	79.0	433	34	AA497734	AA497734 vi72b12.r
26	15.8	79.0	509	34	AA510243	AA510243 v935h09.r
27	15.8	79.0	386	36	AA622988	AA622988 np58h04.s
28	15.8	79.0	607	48	AI614082	AI614082 v935h09.y
29	15.8	79.0	460	48	AI614159	AI614159 vi72b12.y
30	15.8	79.0	424	48	AI614179	AI614179 vi76g02.y
31	15.8	79.0	731	53	HSM007756	AL042906 Homo sapi
32	15.4	77.0	262	27	C03900	C03900 C03900 Huma
33	15.4	77.0	714	29	AA153319	AA153319 ms61c12.r
34	15.4	77.0	463	30	AA210873	AA210873 zr88d07.r
35	15.4	77.0	451	36	C73686	C73686 C73686 Rice
36	15.4	77.0	346	41	AI044171	AI044171 UI-R-Cl-1
37	15.2	76.0	255	20	T08311	T08311 EST06202 In
38	15.2	76.0	330	20	T29096	T29096 EST68411 Hu
39	15.2	76.0	402	21	R07242	R07242 yf14e05.r1
40	15.2	76.0	391	21	RI4146	RI4146 yf78c01.r1
41	15.2	76.0	510	21	T77597	T77597 yc91e03.r1
42	15.2	76.0	526	21	T81887	T81887 yd34g11.r1
43	15.2	76.0	424	21	T85981	T85981 yd62d07.r1
44	15.2	76.0	452	21	T87073	T87073 yd88a09.r1
45	15.2	76.0	590	54	HSM010939	AL046089 Homo sapi

ALIGNMENTS

RESULT 1
F07416 LOCUS F07416 HSC25B041 359 bp mRNA EST 20-FEB-1995
DEFINITION HSC25B041 normalized infant brain cdna Homo sapiens cDNA clone
c-25b04, mRNA sequence.
ACCESSION F07416
NID 9673074
VERSION F07416.1 GI:673074

KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 359)

AUTHORS Auffray, C., Behar, G., Bois, F., Bouchier, C., da Silva, C.,
Dorennes, M.D., Duprat, S., Houligatte, R., Jumeau, M.N., Lamy, B.,
Lorenzo, F., Mitchell, H., Mariage-Samson, R., Pietu, G., Poullot, Y.,
Sebastiani-Kabatchis, C. and Tessier, A.

TITLE IMAGE: molecular integration of the analysis of the human genome
and its expression

JOURNAL C. R. Acad. Sci. III, Sci. Vie 318 (2), 263-272 (1995)
MEDLINE 95277534
COMMENT On Sep 21, 1992 this sequence version replaced gi:276086.

Contact: Genethon
Genexpress-Genethon
Genethon Centre de recherche sur le Genome Humain
1, rue de l'Internationale, BP60 91002 EVRY Cedex, FRANCE
Tel: 33169472800
Fax: 33160778698
Email: genexpress@genethon.fr

Single read.
Genexpress_library_id: C; Genexpress_sequence_id: y1c-25b04
Seq primer: (-21)M13_universal
High quality sequence stop: 415.

FEATURES
source
Location/Qualifiers
1..359
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="c-25b04"
/sex="Female"
/tissue_type="total brain"
/dev_stage="3 months old"
/note="Organ: brain; Vector: lafmid BA; Site_1: HindIII;
Site_2: NotI; sex=Female; dev_stage=3 months old;
isolate=muscular atrophy patient; tissue_type=total
brain; total mRNA was oligo-(dT) primed and directionally
cloned 5' -> 3' into the HindIII -> NotI sites of the
lafmid BA vector. Clone library from B.Souares, Psychiatry
Dept. Columbia University, USA. Normalization_method:
Bento Soares, P.N.A.S in press"

BASE COUNT 102 a 75 c 61 g 120 t 1 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 21; Length 359;
Best Local Similarity 94.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
||||||| |||||||
Db 272 TGGCATTGAACTGGTTCAT 290

RESULT 2
AA214464
LOCUS AA214464 429 bp mRNA EST 12-MAR-1998
DEFINITION IMAGE:649129 5', mRNA sequence.
ACCESSION AA214464
NID g1813283
VERSION AA214464.1 GI:1813283
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 429)

REFERENCE Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.

Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On Nov 29, 1993 this sequence version replaced gi:635188.

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu

This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1360 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 375.

FEATURES
source
Location/Qualifiers
1..429
/organism="Homo sapiens"
/db_xref="GDB:5278408"
/db_xref="taxon:9606"
/clone="IMAGE:649129"
/clone_lib="Stratagene hNT neuron (#937233)"
/dev_stage="hNT neurons"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dT.
Differentially, post mitotic hNT neurons. Average insert
size: 1.5 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTT 3' "

BASE COUNT 120 a 91 c 70 g 147 t 1 others
ORIGIN

Query Match 87.0%; Score 17.4; DB 30; Length 429;
Best Local Similarity 94.7%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCAT 19
||||||| |||||||
Db 315 TGGCATTGAACTGGTTCAT 333

RESULT 3
AI365968/c
LOCUS AI365968 430 bp mRNA EST 07-JAN-1999
DEFINITION ao91c06.x1 Schiller meningioma Homo sapiens cDNA clone
IMAGE:1953226 3' similar to SW:HES1_RAT Q04666 TRANSCRIPTION FACTOR
HES-1. [1] ;, mRNA sequence.

ACCESSION AI365968
NID g4125657
VERSION AI365968.1 GI:4125657
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homidae; Homo.
1 (bases 1 to 430)

REFERENCE Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S.,
Krizman, D., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M.,
Martin, J., Moore, B., Schellenberg, K., Steptoe, M., Tan, F.,
Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.
WashU-NCI human EST Project
Unpublished (1997)
On Jan 17, 1998 this sequence version replaced gi:2044405.

TITLE
JOURNAL
COMMENT

Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

FEATURES

[illegible]

Qy 1 TGGCATTGATCTGGTTCAT 19
|||
Db 116 TGGCATTGATCTGGGTCAT 98

RESULT	4
LOCUS	T27501/c
DEFINITION	T27501 516 bp mRNA EST 12-JUN-1996
ACCESSION	U01116 WATM1 Homo sapiens cDNA clone 27116 similar to close to nucleolin, gb 60858 HUMNUCLEO, mRNA sequence.
NID	T27501
VERSION	9624976
KEYWORDS	T27501.1 GI:624976
SOURCE	EST.
	human.

SOURCE: Humanat.
ORGANISM: Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 516)

AUTHORS	Bouillaud, F.
TITLE	Study of expressed sequences tags in adipose tissue 1994
JOURNAL	Unpublished (1994)
COMMENT	On Sep 21, 1992 this sequence version replaced qi:279106

FEATURES	SOURCE
1. Feature 1	Source 1
2. Feature 2	Source 2
3. Feature 3	Source 3
4. Feature 4	Source 4
5. Feature 5	Source 5
6. Feature 6	Source 6
7. Feature 7	Source 7
8. Feature 8	Source 8
9. Feature 9	Source 9
10. Feature 10	Source 10
11. Feature 11	Source 11
12. Feature 12	Source 12
13. Feature 13	Source 13
14. Feature 14	Source 14
15. Feature 15	Source 15
16. Feature 16	Source 16
17. Feature 17	Source 17
18. Feature 18	Source 18
19. Feature 19	Source 19
20. Feature 20	Source 20
21. Feature 21	Source 21
22. Feature 22	Source 22
23. Feature 23	Source 23
24. Feature 24	Source 24
25. Feature 25	Source 25
26. Feature 26	Source 26
27. Feature 27	Source 27
28. Feature 28	Source 28
29. Feature 29	Source 29
30. Feature 30	Source 30
31. Feature 31	Source 31
32. Feature 32	Source 32
33. Feature 33	Source 33
34. Feature 34	Source 34
35. Feature 35	Source 35
36. Feature 36	Source 36
37. Feature 37	Source 37
38. Feature 38	Source 38
39. Feature 39	Source 39
40. Feature 40	Source 40
41. Feature 41	Source 41
42. Feature 42	Source 42
43. Feature 43	Source 43
44. Feature 44	Source 44
45. Feature 45	Source 45
46. Feature 46	Source 46
47. Feature 47	Source 47
48. Feature 48	Source 48
49. Feature 49	Source 49
50. Feature 50	Source 50
51. Feature 51	Source 51
52. Feature 52	Source 52
53. Feature 53	Source 53
54. Feature 54	Source 54
55. Feature 55	Source 55
56. Feature 56	Source 56
57. Feature 57	Source 57
58. Feature 58	Source 58
59. Feature 59	Source 59
60. Feature 60	Source 60
61. Feature 61	Source 61
62. Feature 62	Source 62
63. Feature 63	Source 63
64. Feature 64	Source 64
65. Feature 65	Source 65
66. Feature 66	Source 66
67. Feature 67	Source 67
68. Feature 68	Source 68
69. Feature 69	Source 69
70. Feature 70	Source 70
71. Feature 71	Source 71
72. Feature 72	Source 72
73. Feature 73	Source 73
74. Feature 74	Source 74
75. Feature 75	Source 75
76. Feature 76	Source 76
77. Feature 77	Source 77
78. Feature 78	Source 78
79. Feature 79	Source 79
80. Feature 80	Source 80
81. Feature 81	Source 81
82. Feature 82	Source 82
83. Feature 83	Source 83
84. Feature 84	Source 84
85. Feature 85	Source 85
86. Feature 86	Source 86
87. Feature 87	Source 87
88. Feature 88	Source 88
89. Feature 89	Source 89
90. Feature 90	Source 90
91. Feature 91	Source 91
92. Feature 92	Source 92
93. Feature 93	Source 93
94. Feature 94	Source 94
95. Feature 95	Source 95
96. Feature 96	Source 96
97. Feature 97	Source 97
98. Feature 98	Source 98
99. Feature 99	Source 99
100. Feature 100	Source 100

```
1. .516
/organism="Homo sapiens"
/strain="caucasian"
```

```
Query Match      84.0%; Score 16.8; DB 20; Length 516;
Best Local Similarity 90.0%; Pred. No. 85;
```

RESULT 5
T34213/C

LOCUS	T34213	296 bp	mRNA	EST	06-SEP-1995
DEFINITION	EST64297	Human White blood cells	Homo sapiens	cdna 5' end similar to None. mRNA sequence.	

ACCESSION	T34213
NID	g616311
VERSION	T34213.1
KEYWORDS	EST.
SOURCE	human.

ORGANISM	Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;	
Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 296)

Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A., Bult, C.J., Lee, N., Kirkness, E.F., Weinstein, K.G., Gockayne, J.D., White, O., Sutton, G., Blake, J.A., Brandon, R.C., Chiu, M.-W., Clayton, R.A., Cline, R.T., Cotton, M.D., Earle-Hughes, J., Fine, L.D., FitzGerald, L.M., FitzHugh, W.M., Fritchman, J.L., Geoghagen, N.S.M., Glodek, A., Gnehm, C.L., Hanna, M.C., Hedblom, E., Hinkle Jr, P.S., Kelley, J.M., Klinek, K.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Merrick, J.W., Moreno-Palauques, R.F., McDonald, L.A., Nguyen, D.T., Pellegriño, S.M., Phillips, C.A., Ryder, S.E., Scott, J.L., Saudek, D.M., Shirley, R., Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y., Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J., Dimke, D., Feng, P., Ferrie, A., Fischer, C., Hastings, G.A., He, W.-W., Hu, J.-S., Greene, J.H., Gruber, J., Hudson, P., Kim, A., Kozak, D.L., Kunsch, C., Ji, H., Li, H., Meisner, P., Olsen, H., Raymond, L., Wei, Y.-F., Wing, J., Xu, C., Yu, G.-L., Ruben, S.M., Dillion, P.J., Fannon, M.R., Rosen, C.A., Haseeltine, W.A., Fields, C., Fraser, C.M. and Venter, J.C.
Initial Assessment of Human Gene Diversity and Expression Patterns Based Upon 83 Million Basepairs of cDNA Sequence
Nature 377, 3-174 (1995)
JOURNAL 96076280
MEDLINE

Contact: Venter, JC
The Institute for Genomic Research
932 Clopper Rd, Gaithersburg, MD 20878
Tel: 3018699056
Fax: 3018699423

Email: tdbinfo@tdb.tigr.org
For clone availability, add
information related to this
(tdbinfo@tdb.tigr.org)
Seq primer: M13 Reverse.

FEATURES	SOURCE
1. High Accuracy: The model achieves a high accuracy rate, indicating its effectiveness in predicting the target variable.	Model Performance Metrics
2. Robustness: The model demonstrates robustness against various input variations and noise, ensuring reliable predictions.	Model Performance Metrics
3. Interpretability: The model's predictions are interpretable, allowing users to understand the underlying factors influencing the outcome.	Model Performance Metrics
4. Scalability: The model is scalable, capable of handling large volumes of data and complex tasks efficiently.	Model Performance Metrics
5. Real-time Processing: The model supports real-time data processing, enabling immediate decision-making and action.	Model Performance Metrics
6. Integration with Existing Systems: The model seamlessly integrates with existing data management and analytics systems.	Model Performance Metrics
7. Customizable Parameters: The model offers customizable parameters to tailor its predictions to specific user requirements.	Model Performance Metrics
8. Regular Updates: The model is regularly updated with the latest data and algorithms to maintain its predictive accuracy.	Model Performance Metrics
9. Compliance with Regulations: The model adheres to relevant data protection and privacy regulations, ensuring ethical data usage.	Model Performance Metrics
10. User-Friendly Interface: The model is accompanied by a user-friendly interface, facilitating easy interaction and data visualization.	Model Performance Metrics

```

1. .296
/organism="Homo sapiens"
/db_xref="ATCC (inhost):104942"
/db_xref="taxon:9606"
/map="5"

```

BASE COUNT		95 a 60 c 81 g 58 t 2 others		115 a 133 c 139 g 107 t			
ORIGIN		/clone_lib="Human white blood cells" /tissue_type="white blood cells" /note="Organ: blood"					
Query Match		84.0%; Score 16.8; DB 20; Length 296;		84.0%; Score 16.8; DB 30; Length 494;			
Best Local Similarity		90.0%; Pred. No. 80;		90.0%; Pred. No. 84;			
Matches		18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;		18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;			
QY	1	TGGCATTGATCTGTTTCATC 20		QY	1	TGGCATTGATCTGTTTCATC 20	
Db	113	TGGCATTGTTCTTGTTCATC 94		Db	473	TGGCATTGAACTGGTTCTTC 492	
RESULT 6		RESULT 7		AA305985		443 bp mRNA	
AA265846		AA265846		EST176975		Jurkat T-cells VI Homo sapiens cDNA 5' end similar to	
LOCUS		LOCUS		EIF-3 p110 subunit, mRNA sequence.			
DEFINITION		DEFINITION		AA305985			
mz53d03.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA		mz53d03.r1 Barstead mouse pooled organs MPLRB4 Mus musculus cDNA		AA305985.1		GI:1958314	
clone IMAGE:717125 5' similar to gb:M1961 CYTOCHROME C OXIDASE		clone IMAGE:717125 5' similar to gb:M1961 CYTOCHROME C OXIDASE		human.			
POLYPEPTIDE VB PRECURSOR (HUMAN); gb:M77040 Mouse cytochrome c		POLYPEPTIDE VB PRECURSOR (HUMAN); gb:M77040 Mouse cytochrome c		Homo sapiens			
oxidase Vb subunit gene, complete cds (MOUSE);, mRNA sequence.		oxidase Vb subunit gene, complete cds (MOUSE);, mRNA sequence.		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AA265846		AA265846		Eutheria; Primates; Catarrhini; Homnidae; Homo.			
NID		NID		1 (bases 1 to 443)			
VERSION		VERSION		Adams, M.D., Kerlavage, A.R., Fleischmann, R.D., Fuldner, R.A.,			
KEYWORDS		KEYWORDS		Bult, C.J., Lee, N.H., Kirkness, E.F., Weinstock, K.G., Gocayne, J.D.,			
SOURCE		SOURCE		White, O., Sutton, G., Blake, J.A., Brandon, R.C., Man-Wai, C.,			
house mouse.		house mouse.		Clayton, R.A., Cline, T.R., Cotton, M.D., Earle-Hughes, J., Fine, L.D.,			
ORGANISM		ORGANISM		Fitzgerald, L.M., Fitzhugh, W.M., Hanna, M.C., Hedblom, E., Hinkle, P.S., Jr.,			
Mus musculus		Mus musculus		Glodek, A., Gnehm, C.L., Liu, L.-I., Marmaros, S.M., Merrick, J.M.,			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		Kelley, J.M., Kelley, J.C., Liu, L.-I., Marmaros, S.M., Peiligrino, S.M.,			
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		Moreno-Palancas, R.F., McDonald, L.A., Nguyen, D.T., Peiligrino, S.M.,			
1 (bases 1 to 494)		1 (bases 1 to 494)		Phillips, C.A., Ryder, S.E., Scott, J.L., Sauder, D.M., Shirley, R.,			
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,		Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,		Small, K.V., Spriggs, T.A., Utterback, T.R., Weidman, J.F., Li, Y.,			
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,		Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,		Bednarek, D.P., Cao, L., Cepeda, M.A., Coleman, T.A., Collins, E.J.,			
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,		Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,		Dimke, D., Feng, D.-F., Ferrie, A., Fischer, C., Hastings, G.A.,			
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and		Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and		He, W.W., Hu, J.S., Greene, J.M., Gruber, J., Hudson, P., Kim, A.K.,			
Waterston, R.		Waterston, R.		Kozak, D.L., Kunsch, C., Hungjun, J., Li, H., Meissner, P.S., Olsen, H.,			
The WashU-HMI Mouse EST Project		The WashU-HMI Mouse EST Project		Raymond, L., Wei, Y.F., Wing, J., Xu, C., Yu, G.L., Ruben, S.M.,			
Unpublished (1996)		Unpublished (1996)		Dillon, P.J., Fannon, M.R., Rosen, C.A., Haseltine, W.A., Fields, C.,			
On Apr 14, 1993 this sequence version replaced gi:638332.		On Apr 14, 1993 this sequence version replaced gi:638332.		Fraser, C.M. and Venter, J.C.			
Contact: Marra M/Mouse EST Project		Contact: Marra M/Mouse EST Project		Initial assessment of human gene diversity and expression patterns			
WashU-HMI Mouse EST Project		WashU-HMI Mouse EST Project		based upon 83 million nucleotides of cDNA sequence			
Washington University School of MedicineP		Washington University School of MedicineP		Nature 377 (6547 Suppl), 3-174 (1995)			
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		96026280			
Tel: 314 286 1800		Tel: 314 286 1800		On Dec 30, 1996 this sequence version replaced gi:1530941.			
Fax: 314 286 1810		Fax: 314 286 1810		Other_ESTs: THC189756			
Email: mouseest@watson.wustl.edu		Email: mouseest@watson.wustl.edu		Contact: Kerlavage, AR			
This clone is available royalty-free through LLNL ; contact the		This clone is available royalty-free through LLNL ; contact the		Bioinformatics			
IMAGE Consortium (info@image.llnl.gov) for further information.		IMAGE Consortium (info@image.llnl.gov) for further information.		The Institute for Genomic Research			
MGI:442621		MGI:442621		9712 Medical Center Drive, Rockville, MD 20850 USA			
Seq primer: -28m13 rev2 ET from Amersham		Seq primer: -28m13 rev2 ET from Amersham		Tel: 3018699056			
High quality sequence stop: 220.		High quality sequence stop: 220.		Fax: 3018699423			
Location/Qualifiers		Location/Qualifiers		Email: arkerlav@tigr.org			
1. .494		1. .494		For clone availability, additional sequence and expression			
/organism="Mus musculus"		/organism="Mus musculus"		Information related to this EST, please check the TIGR Human Gene			
/strain="FVB/N"		/strain="FVB/N"		Index (http://www.tigr.org/tadb/hgi/hgi.html)			
/db_xref="taxon:10090"		/db_xref="taxon:10090"		Seq primer: M13 Reverse.			
/map="5"		/map="5"		Location/Qualifiers			
/clone="IMAGE:717125"		/clone="IMAGE:717125"		1. .443			
/clone_lib="Barstead mouse pooled organs MPLRB4"		/clone_lib="Barstead mouse pooled organs MPLRB4"		/organism="Homo sapiens"			
/sex="mixed"		/sex="mixed"		/db_xref="ArCC (inhst):160079"			
/tissue_type="pooled organs"		/tissue_type="pooled organs"		/db_xref="taxon:9606"			
/dev_stage="7 day"		/dev_stage="7 day"		/clone_lib="Jurkat T-cells VI"			
/lab_host="DH10B"		/lab_host="DH10B"		/cell_type="T-lymphocyte"			
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with		/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with		/note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:			
a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st		a modified polylinker; Site_1: EcoRI; Site_2: NotI; 1st		YhoI"			
strand cDNA was primed with a Not I - oligo(dT) primer [5'		strand cDNA was primed with a Not I - oligo(dT) primer [5'					
TGTTACGATCTGAAGTCGAGCGCCGCTTTTITTTTTTTTTTTTTT		TGTTACGATCTGAAGTCGAGCGCCGCTTTTITTTTTTTTTTTTTT					
3']; double-stranded cDNA was ligated to Eco RI adaptors		3']; double-stranded cDNA was ligated to Eco RI adaptors					
[CTTGATCGGTACC]. digested with Not I and cloned into		[CTTGATCGGTACC]. digested with Not I and cloned into					
the Not I and Eco RI sites of the modified pT73 vector.		the Not I and Eco RI sites of the modified pT73 vector.					
Library constructed by Bob Barstead."		Library constructed by Bob Barstead."					


```

Query Match      84.0%; Score 16.8; DB 51; Length 451;
Best Local Similarity 90.0%; Pred. No. 83;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20
    ||||| |||||
Db 266 TGGCATTGATCAGGTCATC 247

RESULT 13
ID HSM007819 standard: RNA; EST: 867 BP.
XX AC AL042969;
XX SV AL042969.1
XX NI e1401310
XX DT 12-MAR-1999 (Rel. 59, Created)
XX DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX DE Homo sapiens mRNA; EST DKFZp434M1522_r1 (from clone
XX DE DKFZp434M1522)
XX KW EST; expressed sequence tag.
XX OS Homo sapiens (human)
XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
XX OC Primates; Catarrhini; Hominidae; Homo.
XX RN [1]
XX RP 1-867
XX RA Blum H., Bauersachs S., Mewes W., Gassenhuber J., Wiemann S.;
XX RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
XX RL MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX CC Clone from S. Wiemann, sequenced by LMU within the cDNA
XX CC sequencing consortium of the German Genome Project
XX CC No sl sequence available
XX CC This clone is available at the RZPD in Berlin
XX CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
XX CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX FH Key
XX FH Location/Qualifiers
FT source
FT 1. 867
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="DKFZp434M1522"
FT /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT DH10B; sites NotI + SalI"
FT /dev_stage="adult"
FT /tissue_type="testis"
XX SQ Sequence 867 BP; 242 A; 200 C; 237 G; 178 T; 10 other;

Query Match      84.0%; Score 16.8; DB 53; Length 867;
Best Local Similarity 90.0%; Pred. No. 89;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20
    ||||| |||||
Db 527 TGGCATTGTTCTTGTTCATC 508

RESULT 14
ID HSM009381 standard: RNA; EST: 708 BP.
XX ID HSM009381

```

```

AC AL044531;
XX SV AL044531.1
XX NI e1402968
XX DT 12-MAR-1999 (Rel. 59, Created)
XX DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX DE Homo sapiens mRNA; EST DKFZp434K122_r1 (from clone
XX DE DKFZp434K122)
XX KW EST; expressed sequence tag.
XX OS Homo sapiens (human)
XX OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
XX OC Primates; Catarrhini; Hominidae; Homo.
XX RN [1]
XX RP 1-708
XX RA Ansorge W., Wirkner U., Mewes W., Gassenhuber J., Wiemann S.;
XX RT Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
XX RL MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY
XX CC Clone from S. Wiemann, sequenced by EMBL within the cDNA
XX CC sequencing consortium of the German Genome Project
XX CC sl sequence also available
XX CC This clone is available at the RZPD in Berlin
XX CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
XX CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de
XX FH Key
XX FH Location/Qualifiers
FT source
FT 1. 708
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="DKFZp434K122"
FT /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT DH10B; sites NotI + SalI"
FT /dev_stage="adult"
FT /tissue_type="testis"
XX SQ Sequence 708 BP; 241 A; 135 C; 214 G; 118 T; 0 other;

Query Match      84.0%; Score 16.8; DB 54; Length 708;
Best Local Similarity 90.0%; Pred. No. 87;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTCATC 20
    ||||| |||||
Db 168 TGGCATTGTTCTTGTTCATC 149

RESULT 15
ID HSM010310/c
XX AC AL045460;
XX SV AL045460.1
XX NI e1403934
XX DT 12-MAR-1999 (Rel. 59, Created)
XX DT 12-MAR-1999 (Rel. 59, Last updated, Version 1)
XX DE Homo sapiens mRNA; EST DKFZp434G055_r1 (from clone
XX DE DKFZp434G055)
XX KW EST; expressed sequence tag.
XX

```

OS Homo sapiens (human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria;
OC Primates; Catarrhini; Hominidae; Homo.

XX
RN [1]
RP 1-539
RA Duesterhoeft A., Lauber J., Mewes W., Gassenhuber J., Wiemann S.;
FT ;
RL Submitted (12-MAR-1999) to the EMBL/GenBank/DBJ databases.
RL MIPS, Am Klopferspitz 18a D-82152 Martinsried, GERMANY

XX
CC Clone from S. Wiemann, sequenced by Qiagen within the cDNA
CC sequencing consortium of the German Genome Project
CC No s1 sequence available
CC This clone is available at the RZPD in Berlin
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
CC Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

XX
FH Key Location/Qualifiers
FT source 1..539
FT /db_xref="taxon:9606"
FT /organism="Homo sapiens"
FT /clone="DKFZp434G055"
FT /clone_lib="434 (synonym: htes3). Vector pSport1; host
FT DH10B; sites NotI + SalI"
FT /dev_stage="adult"
FT /tissue_type="testis"

XX
SQ Sequence 539 BP; 190 A; 105 C; 163 G; 81 T; 0 other;

Query Match 84.0%; Score 16.8; DB 54; Length 539;
Best Local Similarity 90.0%; Pred. No. 85;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TGGCATTGATCTGGTTTCATC 20
||||||| ||| |||||
DB 30 TGGCATTGTTCTTGTTCATC 11

Search completed: September 18, 1999, 06:47:34
Job time: 3651 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 15:48:25 ; Search time 436.05 Seconds
(without alignments)
145.869 Million cell updates/sec

Title: US-09-037-472-2

Perfect score: 20

Sequence: 1 GTACCTTCGAGTATACATT 20

Scoring table: IDENTITY_NUC

Searched: 679419 seqs, 1590154680 residues

Database : GenEmbl.*

1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_pl1.*
8: gb_pl2.*
9: gb_pr1.*
10: gb_pr2.*
11: gb_pr3.*
12: gb_ro.*
13: gb_st.*
14: gb_sy.*
15: gb_sy.*
16: gb_un.*
17: gb_v1.*
18: em_fun.*
19: em_htg.*
20: em_hum1.*
21: em_hum2.*
22: em_in.*
23: em_om.*
24: em_or.*
25: em_ov.*
26: em_pat.*
27: em_ph.*
28: em_pl.*
29: em_ro.*
30: em_sy.*
31: em_sv.*
32: em_un.*
33: em_v1.*
34: gb_htg1.*
35: gb_htg2.*
36: gb_in1.*
37: gb_in2.*
38: em_ba1.*
39: em_ba2.*
40: em_hum3.*
41: em_hum4.*
42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
C 1	16.8	84.0	267	14	DM170B4T	271130 D. melanoga

RESULT 1

DM170B4T/c

LOCUS

DEFINITION

ACCESSION

NID

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

DM170B4T 267 bp DNA STS 16-APR-1996
D. melanogaster STS determined from European Mapping Project
cosmid, sequence tagged site.
271130
GI:1263650
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea;
Drosophilidae; Drosophila.
1 (bases 1 to 267)
European Drosophila Mapping Consortium.
Direct Submission
Submitted (15-APR-1996) Michael Ashburner, Department of Genetics,
Downing St., Cambridge CB2 3EH, England
STS_name = Dm170B4T
clone_name = 170B4
STS_from_promoter = T7
vector_class = cosmid, Lorist 6
origin_of_clone = Oregon-R

ALIGNMENTS

C 3	15.8	80	J	190815	42	AC007066	AC007066 Homo sapi
C 4	15.8	79.0		89840	8	ATAC004705	AC004705 Arabidops
C 5	15.8	79.0		92524	8	ATAC005398	AC005398 Arabidops
C 6	15.8	79.0		172048	9	HS179N16	295152 Homo sapien
C 7	15.8	79.0		1022	10	HUMVFA28	M25855 Human von W
C 8	15.8	79.0		151750	11	AC003666	AC003666 Homo sapi
C 9	15.8	79.0		197082	11	AC006249	AC006249 Homo sapi
C 10	15.8	79.0		116840	34	AC006576	AC006576 Homo sapi
C 11	15.4	77.0		128000	34	AC004046	AC004046 Homo sapi
C 12	15.4	77.0		173422	35	AC007041	AC007041 Homo sapi
C 13	15.2	76.0		10232	2	AE001062	AE001062 Archaeogl
C 14	15.2	76.0		2468	7	SCYGL050W	272372 S.cerevisia
C 15	15.2	76.0		69817	8	ATAC003096	AC003096 Arabidops
C 16	15.2	76.0		71097	8	F15K9	AC005278 Arabidops
C 17	15.2	76.0		3869	9	HSU23143	U23143 Human mitoc
C 18	15.2	76.0		138443	10	HS179E13	AL035603 Human DNA
C 19	15.2	76.0		130027	11	AC004982	AC004982 Homo sapi
C 20	15.2	76.0		155450	11	AC005951	AC005951 Homo sapi
C 21	15.2	76.0		208278	34	AC004948	AC004948 Homo sapi
C 22	15.2	76.0		173142	34	AC005748	AC005748 Homo sapi
C 23	15.2	76.0		211727	34	CNS0000C	AC006018 Homo sapi
C 24	15.2	76.0		187860	34	HS120G22	AL049838 Homo sapi
C 25	15.2	76.0		31334	36	CEL101B11	AL031847 Homo sapi
C 26	15.2	76.0		34388	36	CET28A8	U80931 Caenorhabdi
C 27	15.2	76.0		174725	42	AC007486	292813 Caenorhabdi
C 28	15.2	75.0		86579	8	ATAC006587	AC006587 Arabidops
C 29	15.2	75.0		30949	36	CEF29G6	278543 Caenorhabdi
C 30	15.2	75.0		187966	42	AC006487	AC006487 Homo sapi
C 31	14.8	74.0		2869	1	AB009410	AB009410 Esccillus
C 32	14.8	74.0		233000	1	AP000003	AP000003 Pyrococcu
C 33	14.8	74.0		13514	1	MTFPIA	X68366 M.thermofo
C 34	14.8	74.0		11014	1	MTFPIA	X68367 M.thermofo
C 35	14.8	74.0		1820	1	MVOHISA	M11218 M.voltae hi
C 36	14.8	74.0		192000	2	AE001438	AE001438 Clostridi
C 37	14.8	74.0		21935	4	AB016081	AB016081 Oryzias l
C 38	14.8	74.0		6886	4	TMIGF11	AF033802 Tillapia m
C 39	14.8	74.0		7886	6	AF125376	AF125376 Bacteriop
C 40	14.8	74.0		12840	7	CHEGRPO	X17191 E.gracilis
C 41	14.8	74.0		5281	7	CHEGRNC	X12890 Euglena gra
C 42	14.8	74.0		143172	7	CLEGGCA	X70810 E.gracilis
C 43	14.8	74.0		92376	8	ATAC006282	AC006282 Arabidops
C 44	14.8	74.0		118595	9	HS288L1	282196 Human DNA s
C 45	14.8	74.0		185516	11	AC004485	AC004485 Homo sapi

in_situ_site_primary = 68B
 BLAST_program = BLASTN
 database_searched = ENBL
 date_of_search = 08-01-1996
 BLAST_program = BLASTX
 database_searched = SWISSPROT
 date_of_search = 15-12-1995.

FEATURES
 source
 1..267
 /organism="Drosophila melanogaster"
 /strain="Oregon-R"
 /db_xref="taxon:7227"
 /clone="170B4"

BASE COUNT
 ORIGIN
 84 a 37 c 68 g 78 t

Query Match 84.0%; Score 16.8; DB 14; Length 267;
 Best Local Similarity 90.0%; Pred. No. 20;
 Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTACCTCCGAGTATACATT 20

||||||| || |||||

Db 233 GTACCTCCGAAGATACATT 214

RESULT 2

AC007066 190815 bp DNA PRI 02-APR-1999
 LOCUS Homo sapiens chromosome 9, clone hRPK.355_O.1, complete sequence.
 DEFINITION AC007066
 ACCESSION AC007066
 NID 94508098
 VERSION AC007066.4 GI:4508098
 KEYWORDS HTG.
 SOURCE human.

ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Homiidae; Homo.

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,

Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArellano, K., Depayre, E., Devon, K., Dewar, K.,

Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,

Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,

Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,

Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,

Marquis, N., McEvan, P., McGurk, A., McKernan, K., McLaughlin, J.,

Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,

Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,

Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,

Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,

Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,

Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

•Direct Submission

Submitted (11-MAR-1999) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 190815)

Birken, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,

Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,

Castle, A., Cerny, J., Colangelo, M., Collins, S., Collymore, A.,

Cooke, P., DeArellano, K., Depayre, E., Devon, K., Dewar, K.,

Donelan, L., Doyle, M., Ferreira, P., FitzHugh, W., Forrest, C.,

Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,

Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,

Karatas, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,

Marquis, N., McEvan, P., McGurk, A., McKernan, K., McLaughlin, J.,

Meldrim, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Niloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Tesfaye, S., Torruella-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.

TITLE

JOURNAL

COMMENT

Submitted (02-APR-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 24, 1999 this sequence version replaced gi:4454611.

All repeats were identified using RepeatMasker: Smit, A.F.A. &

Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html.

FEATURES

source

1..190815

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="hRPK.355_O.1"

/map="9"

/chromosome="9"

repeat_region 156..208

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

repeat_region 633..939

```
repeat_region 1997. .20097 /rpt_family="LIME3"
repeat_region 20098. .20397 /rpt_family="AluSc"
repeat_region 20398. .20340 /rpt_family="LIME3"
repeat_region 21108. .21356 /rpt_family="LIME3"
repeat_region 21357. .21650 /rpt_family="LIME3"
repeat_region 21651. .21836 /rpt_family="AluSx"
repeat_region 22094. .22126 /rpt_family="LIME3"
repeat_region 22131. .22431 /rpt_family="AT_rich"
repeat_region 23215. .23504 /rpt_family="AluSc"
repeat_region 23526. .23547 /rpt_family="AluSg"
repeat_region 24777. .24796 /rpt_family="AT_rich"
repeat_region 24852. .25125 /rpt_family="(A)n"
repeat_region 25126. .25152 /rpt_family="AluSx"
repeat_region 25225. .25502 /rpt_family="(CAAA)n"
repeat_region complement(25743. .25861) /rpt_family="AluJo"
repeat_region 26247. .26459 /rpt_family="MIR"
repeat_region 26602. .26840 /rpt_family="MER58A"
repeat_region 26889. .27077 /rpt_family="AluY"
repeat_region complement(27149. .27280) /rpt_family="MIR"
repeat_region complement(28307. .28606) /rpt_family="AluJb"
repeat_region 29765. .30092 /rpt_family="THE1C"
repeat_region 30093. .30176 /rpt_family="AluSg/x"
repeat_region 30177. .30254 /rpt_family="THE1C"
repeat_region 31203. .31500 /rpt_family="AluJb"
repeat_region complement(32076. .32182) /rpt_family="L2"
repeat_region 32272. .32292 /rpt_family="AT_rich"
repeat_region 32759. .32848 /rpt_family="MIR"
repeat_region 32856. .32879 /rpt_family="(TTG)n"
repeat_region complement(32880. .33159) /rpt_family="AluSg"
repeat_region 33160. .33229 /rpt_family="MIR"
repeat_region 33876. .33909 /rpt_family="AT_rich"
repeat_region complement(35568. .35776) /rpt_family="MER58A"
repeat_region 35781. .36177 /rpt_family="LTR37B"
repeat_region complement(36273. .36557) /rpt_family="AluSg"
repeat_region 36719. .36808 /rpt_family="L1MC5"
repeat_region 36817. .37000 /rpt_family="AT_rich"
repeat_region 37009. .37119
```

```
repeat_region /rpt_family="LIM4"
37120. .37460 /rpt_family="AluSg"
repeat_region 37461. .37749 /rpt_family="AluSx"
repeat_region 37462. .37927 /rpt_family="LIM4"
repeat_region 37928. .38236 /rpt_family="AluSx"
repeat_region 38237. .38841 /rpt_family="LIM4"
```

Query Match 80.0%; Score 16; DB 42; Length 190815;
Best Local Similarity 100.0%; Pred. No. 34;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 CTCCGAGTATACATT 20
|||||

Db 39469 CTCCGAGTATACATT 39484

RESULT 3
ATAC004705/c

LOCUS ATAC004705 89840 bp DNA PLN 09-SEP-1998
DEFINITION Arabidopsis thaliana chromosome II BAC F26C24 genomic sequence,
complete sequence.

ACCESSION AC004705

NID 93252804

VERSION AC004705.1 GI:3252804

KEYWORDS HTG.

SOURCE thale cress.

ORGANISM Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.

REFERENCE 1 (bases 1 to 89840)
AUTHORS Rounsley,S.D., Kaul,S., Lin,X., Ketchum,K.A., Crosby,M.L.,
Brandon,R.C., Sykes,S.M., Mason,T.M., Kerlavage,A.R., Adams,M.D.,
Somerville,C.R. and Venter,J.C.

Arabisopsis thaliana chromosome II BAC F26C24 genomic sequence
Unpublished

2 (bases 1 to 89840)

Rounsley,S.D. and Lin,X.

Direct Submission

Submitted (20-MAY-1998) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA, rounsley@tigr.org

3 (bases 1 to 89840)

Rounsley,S.D.

Direct Submission

Submitted (24-JUN-1998) The Institute for Genomic Research, 9712

Medical Center Dr., Rockville, MD 20850, USA

On Jun 24, 1998 this sequence version replaced gi:3242697.

Address all correspondence to:

Steve Rounsley for Genomic Research

The Institute for Genomic Research

9712 Medical Center Dr,

Rockville, MD 20850,

USA

e-mail: rounsley@tigr.org

BAC clone F26C24 is from Arabidopsis chromosome II and is near the
molecular marker mi398.

The orientation of the sequence is from SP6 to T7 end of the BAC
clone.

Genes were identified by a combination of three methods: Gene
prediction programs including GRAIL (available by anonymous ftp
from arthur.epm.ornl.gov), Genefinder (Phil Green, University of
Washington), Genscan (Chris Burge,
http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene
(http://www.cbs.dtu.dk/netpgene/cbsnetpgene.html), searches of the
complete sequence against a peptide database and the Arabidopsis
EST database at TIGR (http://www.tigr.org/cdb/at/at.html).
Annotated genes are named to indicate the level of evidence for

repeat_region 5592. .5618 /rpt_family="(CAA)n" complement(6613. .6657) /rpt_family="AT-rich" complement(6732. .6761) /rpt_family="AT-rich" 7056. .7088 /rpt_family="AT-rich" complement(7155. .7263) /rpt_family="AT-rich" complement(7371. .7519) /note="exon predicted by xgrail, quality excellent" /gene="T6B13.3" complement(join(<7571. .7686,8006. .8191,9280. .>9331)) complement(<7571. .>9331) /gene="T6B13.3" /note="predicted by genscan" complement(join(7571. .7686,8006. .8191,9280. .9331)) /gene="T6B13.3" /note="hypothetical protein" /codon_start=1 /protein_id="AAC69369.1" /db_xref="GI:3810587" /db_xref="PI:3810587" /translation="MDPYVAKYVWSMKFFDKHADVLYRIGENTLMSCMIHSVEAIIY LFGEYLRRPTRODLKRLRIGELRGLMTGTSIDCKDLSLFATNQEVRKDVRAFQ VLQARFAIVNPLI" repeat_region 10201. .10329 /rpt_family="AT-rich" 10403. .10430 /rpt_family="AT-rich" join(10583. .11027,11107. .11799,11889. .11977,12056. .12125, 12199. .12282,12422. .12498,12566. .12647,12995. .13077, 13170. .13236,13328. .13456,13617. .13726,13904. .14255) /gene="T6B13.4" 10583. .14255 /gene="T6B13.4" join(10724. .11027,11107. .11799,11889. .11977,12056. .12125, 12199. .12282,12422. .12498,12566. .12647,12995. .13077, 13170. .13236,13328. .13456,13617. .13726,13904. .14002) /gene="T6B13.4" /codon_start=1 /product="vacuolar sorting receptor-like protein" /protein_id="AAC69370.1" /db_xref="PID:G3810588" /db_xref="GI:3810588" /translation="MKQLLCYLPWLLLSLVSPFNEARFVVEKNSLVSPTSIRKGT HDSAIGNFGIPQYGGSMAGTVVYPKENQSKREFSDFISFKSQPGALPTFLVDKRGD CFFALKVWNAQKAGASAVLVADNVDEPLITMDTPEEDVSAAKYIENTIPSAIVTKGF GEKLKKAISGDMVNILNDWEAVPHDPDRVEYELMTNSNDECGVKCDMLMEFVKDFK GAAQILEKGGTQFRPHYITWYCPHAFITLSZCKSKCINRGYCAPDEQDFSSGIDG KDYVENLRGLCYKVANETGKFWWHDYIDFQICPKMKRYNKDCAESVKSLSGI DSKRIDCMGDDPADLNPVLKEEQDAQVGKGTGRGDVTILPTLVNRRNYRQKLESA VLKALCSGFEESTEPACISTMETNECLDNGGCWQDKSANTITACKDTPRGKVCVP IVGVRFKGDGYSHCGSPGRCITINNGGHEERDGHAFACVDKDSVKCECPGPKF GDGVKEDINECKEKACQCEPCCKNTWGSCECSGDLLYMRDHTCISKTGSQV KSAWAAVWLIMLSLGLAAGAYLVYKRLRYQYNDSEIRAIMAQYMPDLSQPEVPNTN DERA" misc_feature 14649. .14675 /note="exon predicted by xgrail, quality marginal_shadowexon" complement(<14856. .>15944) /gene="T6B13.5" complement(<14856. .>15944) /gene="T6B13.5" /note="predicted by genscan and genefinder" complement(14856. .15944) /gene="T6B13.5" /note="hypothetical protein" /codon_start=1 Query Match 79.08; Score 15.8; DB 8; Length 92524; Best local Similarity 89.58; Pred. No. 47;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0; Qy 2 TACCTTCGAGTATACATT 20 ||||| ||||| ||||| Db 6080 TACCTTCGAGATATATT 6062 RESULT 5 HSI179N16 172048 bp DNA PRI 03-AUG-1998 LOCUS HSI179N16/c Homo sapiens DNA sequence from PAC 179N16 on chromosome 6p21.1-21.33. Contains the SAPK4 (MAPK p38delta) gene, and the alternatively spliced SAPK2 gene coding for CSaids binding protein CSBP2 and a MAPK p38beta like protein. Contains ESTs. STSS and two predicted CpG islands, complete sequence. 295152 ACCESSION 295152.1 GI:3036773 NID 295152.1 GI:3036773 VERSION HTG: Cpg island; CSaids binding protein; CSBP2; MAPK; Mitogen activated protein kinase; MX12; P38; SAPK4; Stress activated protein kinase. human. SOURCE ORGANISM Homo sapiens Eukaryota; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 172048) Palmar,S. Direct Submission Submitted (02-APR-1998) Chromosome 6 Project Group (http://www.sanger.ac.uk/HGP/Chr6/) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Apr 8, 1998 this sequence version replaced gi:2505919. IMPORTANT: This sequence is the entire insert of clone 179N16. During sequence assembly data are compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variations annotated may not be found in the sequence submission corresponding to the overlapping clone as we submit sequences with only a small overlap as described above. This sequence was generated from part of bacterial clone contigs of human chromosome 6, constructed in collaboration by the Sanger Centre chromosome 6 mapping group and Ioannis Ragousis. Further information can be found at http://www.sanger.ac.uk/HGP/Chr6/ This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key. The true left end of clone 179N16 is at 1 in this sequence. The true right end of clone 524E15 is at 690. The true left end of clone 179N16 is at 172048. 179N16 is from the library RPCII constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see http://bacpac.med.buffalo.edu/. FEATURES Location/Qualifiers 1. .172048 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="6" /map="p21.1-21.33" /clone="179N16" /clone_lib="RPCII" complement(1. .1077) /note="match: STS AL020972" 4. .457 /note="L1 repeat: matches 5082. .4624 of consensus" repeat_region 493. .734 /note="AluJo repeat: matches 302. .62 of consensus; incomplete repeat" repeat_region 1714. .1954 /note="L1MB8 repeat: matches 324. .574 of consensus"


```

RESULT 6
HUMVWF28/c
LOCUS HUMVWF28 1022 bp DNA PRI 18-SEP-1997
DEFINITION Human von Willebrand factor gene, exon 35.
ACCESSION M25855 M25716
NID 9340344
VERSION M25855.1 GI:340344
KEYWORDS coagulation factor VIII; von Willebrand factor.
SEGMENT 28 of 38
SOURCE Human placenta, leukocyte, fetal liver and cell line EBI9 DNA
clones lambda- [1,2,7,9,11], h[6,7,18], c[1,2,3,4,5,9,14,18] and
pWE[5,8,123].

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 1022)
AUTHORS Mancuso,D.J., Tuley,E.A., Westfield,L.A., Worrall,N.K.,
Shelton-Inloes,B.B., Sorace,J.M., Alevy,Y.G. and Sadler,J.E.
TITLE Structure of the gene for human von Willebrand factor
JOURNAL J. Biol. Chem. 264 (33), 19514-19527 (1989)
MEDLINE 90062044
COMMENT Draft entry and computer-readable sequence for [1] kindly provided
by J.E.Sadler, 30-JUN-1989.
FEATURES
Source
1..1022
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="12pter-p12"
<1..269
/gene="VWF"
/number=34
exon 270..490
/gene="VWF"
/number=35
intron 491..>1022
/gene="VWF"
/number=35

BASE COUNT 272 a 207 c 238 g 305 t
ORIGIN About 16 kb after segment 27; chromosome 12pter-12.

Query Match 79.0%; Score 15.8; DB 10; Length 1022;
Best Local Similarity 89.5%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCTTCGAGTATACATT 20
|||||
Db 802 TACCTTCGAGTAAACAAT 784

RESULT 7
AC003666/c
LOCUS AC003666 151750 bp DNA PRI 05-MAY-1998
DEFINITION Homo sapiens Xp22 BAC GS-551019 (Genome Systems Human BAC library)
and cosmids U199A7 and U209F2 (Lawrence Livermore X chromosome
cosmid library) containing part of human chloride channel 4 gene,
complete sequence.
ACCESSION AC003666 AC002358 AC002360 AC003017
NID 92982476
VERSION 92982476
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 151750)
AUTHORS Muzny,D., Aronson,A.D., Brundage,E., Carvelli,K., Chen,E., Chen,J.,
Di,W., Ding,Y., Dugan,S., Durbin,J., Forcum,J., Ganesha,R.,
Garcia,C., Goodman,M., Gortrell,J.H., Haywood,M., Jackson,L.,
Jin,S., Kampal,R., Karpathy,S., Leal,B., Li,Y., Liu,W., Logan,O.,
Lu,J., Ly,T., Martinez,C., Oswal,G., Perez,L., Rashid,N.D.,
Rowland,K., Savage,L., Scherer,S.S., Shen,H., Simms,K.M., Todd,J.,
Vo,Q., Worley,K.C., Yu,W., Chinault,C., Nelson,D. and Gibbs,R.A.

```

```

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151750)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (09-DEC-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 151750)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (27-MAR-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.
The repeat regions shown were identified using RepeatMasker by
Adrian Smit.
Sequence similarities were identified using Powerblast by Jinghui
Zhang.
Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.
Location/Qualifiers
1..151750
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GS-551019, U199A7, U209F2"
/chromosome="X"
/clone_lib="Genome Systems Human BAC library, Lawrence
Livermore X chromosome library"
/map="Xp22"
complement(2657..2689)
/rpt_family="AT-rich"
repeat_region
complement(2773..2863)
repeat_region
/rpt_family="MER5B"
3472..4871
repeat_region
/rpt_family="L1PA2"
5843..6143
repeat_region
/rpt_family="L1MA8"
complement(6989..7493)
repeat_region
/rpt_family="L1MA3"
complement(7513..7811)
repeat_region
/rpt_family="AlusX"
11128..11201
repeat_region
/rpt_family="(GGAA)n"
11487..11618
STS
/standard_name="hz-1"
11757..11981
repeat_region
/rpt_family="MER20"
12660..12782
repeat_region
/rpt_family="L2"
12879..13195
repeat_region
/rpt_family="MLTIE"
complement(15364..15659)
repeat_region
/rpt_family="AluJo"
complement(15836..16099)
repeat_region
/rpt_family="MER20"
complement(16717..17085)
repeat_region
/rpt_family="L1MB7"
complement(17159..17374)
repeat_region
/rpt_family="L1MB7"

```



```
repeat_region complement(17421..19544)
repeat_region /rpt_family="LIM4"
repeat_region complement(19654..19760)
repeat_region /rpt_family="LMB7"
repeat_region complement(20189..20669)
repeat_region /rpt_family="LMB7"
repeat_region 20670..21171
repeat_region /rpt_family="MLT1F"
repeat_region complement(21172..21792)
repeat_region /rpt_family="LMB7"
repeat_region 22005..22430
repeat_region /rpt_family="LIMC1"
repeat_region 23332..23445
repeat_region /rpt_family="LIP4A"
repeat_region 23469..23588
repeat_region /rpt_family="FLAM_A"
repeat_region complement(25042..25067)
repeat_region /rpt_family="POLY_A"
repeat_region complement(25163..25265)
repeat_region /rpt_family="LIPAS"
repeat_region 25264..25431
repeat_region /rpt_family="LIPAS"
repeat_region complement(25569..26063)
repeat_region /rpt_family="MER31A"
repeat_region complement(27469..27570)
repeat_region /rpt_family="MIR"
repeat_region 28608..28643
repeat_region /rpt_family="(TAAAA)n"
repeat_region complement(28713..29010)
repeat_region /rpt_family="AluJb"
repeat_region 29209..29310
repeat_region /rpt_family="MIR"
repeat_region 29312..29613
repeat_region /rpt_family="AluSx"
repeat_region 30127..30683
repeat_region /rpt_family="MLT1D"
repeat_region 31072..31160
repeat_region /rpt_family="(GAA)n"
repeat_region 31161..31279
repeat_region /rpt_family="(GGA)n"
repeat_region complement(31772..31891)
repeat_region /rpt_family="(GA)n"
repeat_region 32352..32572
repeat_region /rpt_family="MIR"
repeat_region complement(33455..33479)
repeat_region /rpt_family="AT-rich"
repeat_region 34912..35382
repeat_region /rpt_family="L1MA1"
repeat_region 35405..35701
repeat_region /rpt_family="AluSg"
repeat_region 35707..36176
repeat_region /rpt_family="L1MA1"
repeat_region 36178..38428
repeat_region /rpt_family="L1PA2"
STS 36898..37069
/standard_name="STS1-cSRL-31b6-uA/cSRL-31b6-u2, Chr. -,
Homo sapiens"
/db_xref="dbSTS:7821"
37958..38089
/standard_name="A002D07, Chr. -, Homo sapiens"
/db_xref="dbSTS:32826"
38436..38566
repeat_region /rpt_family="L1MA1"
38597..38795
repeat_region /rpt_family="L1PB3"
repeat_region complement(38769..40304)
repeat_region /rpt_family="L1PA12"
40309..40609
repeat_region /rpt_family="AluY"
repeat_region complement(40626..40905)
repeat_region /rpt_family="L1PA12"
40904..41028
repeat_region /rpt_family="L1PB3"
```

```
repeat_region complement(41977..42037)
repeat_region /rpt_family="AT-rich"
repeat_region complement(42391..42718)
repeat_region /rpt_family="MER34"
repeat_region 42979..43067
repeat_region /rpt_family="L2"
repeat_region complement(43085..43220)
repeat_region /rpt_family="MIR"
repeat_region 43446..43746
repeat_region /rpt_family="AluSx"
repeat_region 44344..44380
repeat_region /rpt_family="(CAAAA)n"
repeat_region complement(45352..45584)
repeat_region /rpt_family="MIR"
repeat_region complement(46643..46707)
repeat_region /rpt_family="L2"
repeat_region complement(46905..47293)
repeat_region /rpt_family="MSTA"
repeat_region complement(48279..48405)
repeat_region /rpt_family="MERSA"
repeat_region 48406..48701
repeat_region /rpt_family="AluSx"
repeat_region complement(48710..48776)
repeat_region /rpt_family="MERSA"
repeat_region complement(49213..49456)
repeat_region /rpt_family="L1MB3"
repeat_region complement(50681..50736)
repeat_region /rpt_family="MIR"
repeat_region complement(50751..50796)
repeat_region /rpt_family="L2"
repeat_region 51052..51443
repeat_region /rpt_family="MLT1A2"
repeat_region 51480..51927

Query Match 79.0%; Score 15.8; DB 11; Length 151750;
Best Local Similarity 89.5%; Pred. No. 45;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTACCTTCGAGTATACAT 19
||||||| ||||| |||
Db 82734 GTACCTTCAGATATGCAT 82716

RESULT 8
AC006249/c
LOCUS AC006249 197082 bp DNA PRI 25-DEC-1998
DEFINITION Homo sapiens chromosome 18, clone hRPK.430_E_17, complete sequence.
ACCESSION AC006249
NID 94062902
VERSION AC006249.1 GI:4062902
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 197082)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 18, clone hRPK.430_E_17
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 197082)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,
Castle,A., Cerny,J., Collangelo,M., Collins,S., Collymore,A.,
Cooke,P., DeArellano,K., Depayre,E., Devon,K., Devay,K.,
Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R.,
Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A.,
Herena,L., Horton,L., Howland,J.C., Jones,C., Kann,L., Karatas,A.,
Lehoczky,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A.,
McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J.,
Mychaleckyj,J., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,
Pavlin,B., Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A.,
Severy,P., Stange-Thomann,N., Stojanovic,N., Stone,C.,
Subramanian,A., Tesfaye,S., Torruella-Miller,I., Vassiliev,H.,
```



```

repeat_region complement(39999..40162)
repeat_region /rpt_family="MLT1-INTERNAL"
40189..40639
repeat_region /rpt_family="MLT1C"
complement(40666..41225)
repeat_region /rpt_family="MLT1E"
complement(41227..41638)
repeat_region /rpt_family="LIMEC"
complement(41761..41856)
repeat_region /rpt_family="LIM4C"
42163..42487
repeat_region /rpt_family="MER7A"
43491..43558
repeat_region /rpt_family="AT_rich"
44384..44408
repeat_region /rpt_family="AT_rich"
complement(44409..44693)
repeat_region /rpt_family="LIPA7"
45397..45710
repeat_region /rpt_family="AluYa8"
45900..46800
repeat_region /rpt_family="Tigger3b"
complement(46803..47354)
repeat_region /rpt_family="MLI2D"
47355..47662
repeat_region /rpt_family="Tigger3b"

Query Match 79.0%; Score 15.8; DB 11; Length 197082;
Best Local Similarity 89.5%; Pred. No. 44;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TACCTTCGAGTATACATT 20
||||||| |||||||
Db 63403 TACCTTCGAGTATACATT 63385

RESULT 9
AC006576 AC006576 116840 bp DNA PRI 14-MAY-1999
LOCUS Homo sapiens Chromosome 22q11.2 BAC Clone 762c12 In CES Region,
DEFINITION complete sequence.
ACCESSION AC006576
NID 94827300
VERSION AC006576.15 GI:4827300
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 116840)
AUTHORS Cantu,L.A. and Roe,B.A.
TITLE Homo sapiens Chromosome 22q11.2 BAC Clone 762c12 In CES Region
JOURNAL Unpublished (1997)
REFERENCE 2 (bases 1 to 116840)
AUTHORS McDermid,H.E.
JOURNAL Unpublished (1997)
REFERENCE 3 (bases 1 to 116840)
AUTHORS Simon,M.I.
JOURNAL Unpublished (1997)
REFERENCE 4 (bases 1 to 116840)
AUTHORS Cantu,L.A., McDermid,H.E., Simon,M.I. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (16-FEB-1999), Department Of Chemistry And Biochemistry,
The University Of Oklahoma., 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE 5 (bases 1 to 116840)
AUTHORS Cantu,L.A., McDermid,H.E., Simon,M.I. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (14-MAY-1999), Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT On May 14, 1999 this sequence version replaced gi:4803972.
FEATURES
Location/Qualifiers
source

source 1..116840
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="22"
/map="22q11.2"
/clone="762c12"

BASE COUNT 34210 a 28605 c 26483 g 27542 t
ORIGIN

Query Match 79.0%; Score 15.8; DB 42; Length 116840;
Best Local Similarity 89.5%; Pred. No. 46;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 TACCTTCGAGTATACATT 20
||||||| |||||||
Db 32331 TACCTTCGAGTAAACAAT 32349

RESULT 10
AC004046 AC004046 128000 bp DNA HTG 16-MAY-1998
LOCUS Homo sapiens chromosome 4, WORKING DRAFT SEQUENCE, 10 unordered
DEFINITION pieces.
ACCESSION AC004046
NID G3135247
VERSION AC004046.1 GI:3135247
KEYWORDS HTG: HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 128000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 128000)
AUTHORS Stone,N.E., Schmutz,J.J., Cox,D.R. and Myers,R.M.
TITLE Direct Submission
JOURNAL Submitted (28-JAN-1998) Department of Genetics, Stanford Human
Genome Center, 855 California Avenue, Palo Alto, CA 94304, USA
On May 16, 1998 this sequence version replaced gi:2979537.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 10 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1158: contig of 1158 bp in length
* 1159 1349: gap of unknown length
* 1350 2553: contig of 1204 bp in length
* 2554 2744: gap of unknown length
* 2745 5805: contig of 3061 bp in length
* 5806 5996: gap of unknown length
* 5997 9107: contig of 3111 bp in length
* 9108 9298: gap of unknown length
* 9299 17597: contig of 8299 bp in length
* 17598 17788: gap of unknown length
* 17789 31282: contig of 13494 bp in length
* 31283 31473: gap of unknown length
* 31474 45223: contig of 13750 bp in length
* 45224 45413: gap of unknown length
* 45414 61363: contig of 15950 bp in length
* 61364 61553: gap of unknown length
* 61554 86512: contig of 24959 bp in length
* 86513 86702: gap of unknown length
* 86703 128000: contig of 41298 bp in length.
* Location/Qualifiers
* 1..128000
/organism="Homo sapiens"
/db_xref="taxon:9606"

FEATURES
source

```

```

BASE COUNT      41094 a 24409 c 23279 g 37501 t 1717 others
ORIGIN
/chromosome="4"

Query Match      77.0%; Score 15.4; DB 34; Length 128000;
Best Local Similarity 94.1%; Pred. No. 77;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 COTTCGAGTATACATT 20
||||| |||||||
Db 73734 COTTCGAGTATACATT 73718

RESULT 11
AC007041/c
LOCUS      AC007041      173422 bp      DNA      HTG      05-JUN-1999
DEFINITION Homo sapiens clone NH0327N17, WORKING DRAFT SEQUENCE, 2 unordered
pieces.
ACCESSION  AC007041
NID        95001509
VERSION    AC007041.2 GI:5001509
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE     human.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 173422)
AUTHORS   Waterston,R.H.
TITLE     The sequence of Homo sapiens clone
JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 173422)
AUTHORS   Waterston,R.H.
TITLE     Direct Submission
JOURNAL   Submitted (06-MAR-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT    On Jun 5, 1999 this sequence version replaced gi:4371304.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 70598: contig of 70598 bp in length
* 70599 70616: gap of unknown length
* 70617 173422: contig of 102806 bp in length.
FEATURES             source
     source          1..173422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="NH0327N17"
BASE COUNT      53415 a 35597 c 34493 g 49899 t 18 others
ORIGIN

Query Match      77.0%; Score 15.4; DB 35; Length 173422;
Best Local Similarity 94.1%; Pred. No. 75;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 4 COTTCGAGTATACATT 20
||||| |||||||
Db 74133 COTTCGAGTATACATT 74117

RESULT 12
AE001062/c
LOCUS      AE001062      10232 bp      DNA      BCT      18-MAY-1998
DEFINITION Archaeoglobus fulgidus section 45 of 172 of the complete genome.
ACCESSION  AE001062 AE000782
NID        92689385

VERSION        AE001062.1 GI:2689385
KEYWORDS       Archaeoglobus fulgidus.
SOURCE         Archaeoglobus fulgidus.
ORGANISM       Archaeoglobus fulgidus.
REFERENCE      1 (bases 1 to 10232)
AUTHORS        Klenk,H.P., Clayton,R.A., Tomb,J., White,O., Nelson,K.E.,
                Ketchum,K.A., Dodson,R.J., Gwinn,M., Hickey,E.K., Peterson,J.D.,
                Richardson,D.L., Kerlavage,A.R., Graham,D.E., Kyripides,N.C.,
                Fleischmann,R.D., Quackenbush,J., Lee,N.H., Sutton,G.G., Gill,S.,
                Kirkness,E.F., Dougherty,B.A., McKenney,K., Adams,M.D., Loftus,B.,
                Peterson,S., Reich,C.I., McNeil,L.K., Badger,J.H., Glodek,A.,
                Zhou,L., Overbeek,R., Gocayne,J.D., Weidman,J.F., McDonald,L.,
                Utterback,T., Cotton,M.D., Spriggs,T., Artiach,P., Kaine,B.P.,
                Sykes,S.M., Sadow,P.W., D'Andrea,K.P., Bowman,C., Fujii,C.,
                Garland,S.A., Mason,T.M., Olsen,G.J., Fraser,C.M., Smith,H.O.,
                Woese,C.R. and Venter,J.C.
TITLE          The complete genome sequence of the hyperthermophilic,
                sulphate-reducing archaeon Archaeoglobus fulgidus
JOURNAL        Nature 390 (6658), 364-370 (1997)
MEDLINE        98049343
REFERENCE      2 (bases 1 to 10232)
AUTHORS        Klenk,H.P., Clayton,R.A., Tomb,J., White,O., Nelson,K.E.,
                Ketchum,K.A., Dodson,R.J., Gwinn,M., Hickey,E.K., Peterson,J.D.,
                Richardson,D.L., Kerlavage,A.R., Graham,D.E., Kyripides,N.C.,
                Fleischmann,R.D., Quackenbush,J., Lee,N.H., Sutton,G.G., Gill,S.,
                Kirkness,E.F., Dougherty,B.A., McKenney,K., Adams,M.D., Loftus,B.,
                Peterson,S., Reich,C.I., McNeil,L.K., Badger,J.H., Glodek,A.,
                Zhou,L., Overbeek,R., Gocayne,J.D., Weidman,J.F., McDonald,L.,
                Utterback,T., Cotton,M.D., Spriggs,T., Artiach,P., Kaine,B.P.,
                Sykes,S.M., Sadow,P.W., D'Andrea,K.P., Bowman,C., Fujii,C.,
                Garland,S.A., Mason,T.M., Olsen,G.J., Fraser,C.M., Smith,H.O.,
                Woese,C.R. and Venter,J.C.
TITLE          Submitted (15-DEC-1997) The Institute for Genomic Research, 9712
                Medical Center Dr, Rockville, MD 20850, USA
JOURNAL        In order to show the genes in ascending order on the genome, the
                origin of this version has been moved by TIGR to position 2093570
                of the original version and the opposite strand is shown from the
                original version.
REMARK         On Dec 16, 1997 this sequence version replaced gi:2649999.
COMMENT        Location/Qualifiers
FEATURES             source
     source          1..10232
                     /organism="Archaeoglobus fulgidus"
                     /db_xref="taxon:2234"
                     complement(254..481)
                     /gene="AF0610"
                     complement(254..481)
                     /gene="AF0510"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="A. fulgidus predicted coding region AF0610"
                     /protein_id="AAB90631.1"
                     /db_xref="PID:g2650012"
                     /db_xref="GI:2650012"
                     /translation="MVVADNLVIERVKKNIEENITEVTFVEFPVINYKKFHKGV
                     LITERGDVLLSIELQRRLRIVRKNHFQIF"
                     /length=1640
                     /gene="AF0611"
                     /length=1640
                     /note="AF0611"
                     /note="similar to SP:P40588 SP:P47187 PID:1015923 percent
                     identity: 22.33; identified by sequence similarity;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAB90626.1"
                     /db_xref="PID:g2650006"
                     /db_xref="GI:2650006"

```

```
/translation="MKHKIKLKGFSLSNEKDIIDFHPWVKPLLEEVNRGNWYFSDV
KAEVLDELKLDLRYPPLEFEGGTVEISAEGSEPPAVLKLVSIESFVR
VSTKCNWAAEIDPFKREVNSIKDLVWAFGEVDKLSQAREVVEVARWLTIEKGFKAN
NVYIKDYKKLVDMFEPKFAVTLTAVEDENKVPWGEELKELSFEFFERGTFGAE
NGSV"
complement(1679..2065)
/gene="AF0612"
CDS
complement(1679..2065)
/gene="AF0612"
/notes="similar to GP:780458 percent identity: 32.20;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB90625.1"
/db_xref="PID:g2650005"
/db_xref="GI:2650005"
/translation="MFVNGTPTVIEVLSNPLLSVKPFESGSLAVIDTGYGFAVVPQ
DIFKRLDELQSKKALTLPTERLIESTGSYARIIPELKTRDGFVETDGVDEIV
LGTFFLEGFKLVLDYCTRFEISSCW"
complement(2055..2372)
/gene="AF0613"
CDS
complement(2055..2372)
/gene="AF0613"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF0613"
/protein_id="AAB90630.1"
/db_xref="PID:g2650011"
/db_xref="GI:2650011"
/translation="MCTLPIISDDVEKRLRDVVKEKHSKGMKSVIEALKIYFS
ILSKKKVFRAYGSELVAEADLEELAKILREKNVDPRSIVKIVSSEHKIPVARMGWK
YVR"
complement(2718..2999)
/gene="AF0614"
CDS
complement(2718..2999)
/gene="AF0614"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF0614"
/protein_id="AAB90629.1"
/db_xref="PID:g2650010"
/db_xref="GI:2650010"
/translation="MKEIKEITKKQVDAEIIYLYGSVVEGDYSIGLSIDIDVAIVSDVF
EDNRKLEFFGKITKKFDSPPFEFHILTKKWKMSKRFIRKRRDLIT"
complement(3104..3493)
/gene="AF0615"
CDS
complement(3104..3493)
/gene="AF0615"
/notes="hypothetical protein; identified by GeneMark;
putative"
/codon_start=1
/transl_table=11
/product="A. fulgidus predicted coding region AF0615"
/protein_id="AAB90627.1"
/db_xref="PID:g2650008"
/db_xref="GI:2650008"
/translation="MTEFLRLKFASGAFGALINTLILYIFTEYFRVYIYSSILAI
EIAIVQFILNDWRTRDKRSSDLRLFLRLKSLNRVAGIFVNISILYVLTLEFKI
YIIFSVGIGCAFIFYNIMESKLTWR"
complement(3490..5091)
/gene="AF0616"
CDS
complement(3490..5091)
/gene="AF0616"
/notes="similar to PID:1652163 percent identity: 28.15;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
```

```
/protein_id="AAB90624.1"
/db_xref="PID:g2650004"
/db_xref="GI:2650004"
/translation="MEVKYLKSEFLVLLITTFVITNVTYFIAMDDSPQSDPALHL
TYSVFNLIKSLQSEIVKVSNIYFFFLYSTPLYFUFSEDITAILNLIIYFIL
IYSVQIAIKMDKNAGVISAVISAFFALIKFORVYIMDFALLSTAFAILYLRSE
DFRNKAYSVLFGIVGIAELTKWNAFYIIPISFALFLVNYFMRCPYCHOIVKNGKV
GYRRCSSKKHLKLYEKSAKNTNMLNFAISLFAFLFAAWWYLPNLSVTIMRLTYFAN
IGGKEDPAFLTQCIWIYANSILDAGVAGVLLFAVSIYLYKNNRYLEGLLTPIA
LIITLILSNKMDPRYIMPVLPFAISGVFVSSLANRKYSTITLIFILLGLLNISA
LTFQPDIDNKILPNPEHKPKDKYALAEIKESGEGKIVVLPDHPYLLNGQSLN
FYRLKEGYFAIYGVYIGYEAFAVHNFDKITYIILIEPREHKGVGYDIERKLYEFFYE
RKDNFEVVKVFDLPDMSKLVYKRT"
complement(5115..6191)
/gene="AF0617"
CDS
complement(5115..6191)
/gene="AF0617"
/notes="similar to GB:L77117 PID:I500505 percent identity:
29.04; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="LPS biosynthesis protein, putative"
/protein_id="AAB90623.1"
/db_xref="PID:g2650003"
/db_xref="GI:2650003"
/translation="MKIAQVCPREYPHIGVETHVYIEIASRIAKKFDVLTDPGCK
LPKVEIDGLTVRRFKSLAPSEAYISPELYDLKKNSSDYDVVHANYHAFFALFAA
FKVKNKLIFTPHYHGSHGSHFFNLVHKPKIFGRKIFKRAADIVCVSNYEKNLKN
FKVAEDRTYIPNGINLDEFKDIKIKRNKESMKKTLILYGRVKEKGLDYVVKSLKH
LPDNFTLEVVGKSGYSKIVEMAKKLDVIDIRIFYODLSRPKELIDRYAKADVLVLSK
HEAYGIVAEALAAKTPCIVANTALSSEIDNKNVFGIDYPIVNSELARLIERVSNVK
AGDKMLDDDDVNERLIRISAKI"
complement(6298..6588)
/gene="AF0618"
CDS
complement(6298..6588)
/gene="AF0618"
/notes="similar to GP:1707702 percent identity: 34.57;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAB90622.1"
/db_xref="PID:g2650002"
/db_xref="GI:2650002"
/translation="MEKNLYDTNKLIELYKKNESMSGYTTILNVIEFPKALEFNLTLV
YLSKSDYRLAIKISTELLKIGEPVAVYTLISAANKQRFKGCNAGTFFAS"
complement(6678..6788)
/gene="AF0619"
CDS
complement(6678..6788)
/gene="AF0619"
```

```
Query Match 76.0%; Score 15.2; DB 2; Length 10232;
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTACCTTCGAGTATACATT 20
||||| | |||||
Db 6387 GTACCTGCTGTATACATT 6368

RESULT 13
SCYGL050W/c
LOCUS SCYGL050W 2468 bp DNA PLN 06-AUG-1997
DEFINITION S.cerevisiae chromosome VII reading frame ORF YGL050w.
ACCESSION 272572 Y13135
NID g1322542
VERSION 272572.1 GI:1322542
KEYWORDS baker's yeast.
SOURCE Saccharomyces cerevisiae
ORGANISM Eukaryota; Fungi; Ascomycota; Hemiascomycetes; Saccharomycetales;
Saccharomycetaceae; Saccharomyces.
REFERENCE 1 (bases 1 to 2468)
```

AUTHORS

MIPS.
Direct Submission
Submitted (14-MAY-1996) Data collected by MIPS on behalf of the European yeast chromosome VII sequencing project. MIPS at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a D-82152 Martinsried, FRG; E-mail: Mewes@mips.emblnet.org

REFERENCE

2 (bases 1 to 2468)
Feuermann, M., de Montigny, J., Potier, S. and Souciet, J.L.
The characterization of two new clusters of duplicated genes suggests a 'lego' organization of the yeast Saccharomyces cerevisiae chromosomes

AUTHORS

Yeast 13 (9), 861-869 (1997)
9/37/993

MEDLINE

9/37/993

FEATURES

Location/Qualifiers

source*

1. .2468 /organism="Saccharomyces cerevisiae"

/db_xref="taxon:4932"

/chromosome="VII"

613..754 /note="delta-remnant"

755..816 /note="delta-remnant"

complement(837..1062)

/note="sigma-remnant"

1079..1150

/gene="tr(TCT)GL - systematic name"

/note="tr(TCT)GL - common name; anticodon gene: TCT"

1079..1150

/gene="tr(TCT)GL - systematic name"

1385..2206 /note="ORF YGL050w"

/codon_start=1

/protein_id="CAA96752.1"

/db_xref="PID:e243282"

/db_xref="PID:g1322543"

/db_xref="GI:1322543"

/db_xref="SWISS-PROT:P53177"

/translation="MAQNAFEQKRAILNEIDSTODLSPKGTIDELGIPITIDLINA

SADMTTSSCGSRVSELETKSYNGEVIGKGGCKWLYVTHDEKVIWGLDELKS

KSEFSELSKEIPTKTVGSIRYILKYEPFLHVKCRDFAASKLYNVMSCGPRE

SGIGNNLAIRINIKLDVLTSGTLKFFVTPYVSVLDSLSLKFENTRKM

QALYDREKELINCAPDVNSKVNITPIETKEERRERKREGMERQRLKSPQNVL

825 a 408 c 488 g 747 t

BASE COUNT

ORIGIN

Query Match 76.08; Score 15.2; DB 7; Length 2468;

Best Local Similarity 85.08; Pred. No. 1.4e+02;

Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTACCTCCGAGTATACATT 20

DB 1578 GTACCTCCAGATACACT 1559

||||||| || ||||| |

RESULT 14

ATAC003096

LOCUS

DEFINITION

ATAC003096 69817 bp DNA PLN 25-SEP-1998

Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence,

complete sequence.

ACCESSION

AC003096

NID

Q3132469

KEYWORDS

HTG

SOURCE

thale cress.

ORGANISM

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core

eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;

Arabidopsis.

1 (bases 1 to 69817)

Rounsley, S.D., Lin, X., Ketchum, K.A., Crosby, M.L., Brandon, R.C.,

Sykes, S.M., Kaul, S., Mason, T.M., Kerlavage, A.R., Adams, M.D.,

et al.

REFERENCE

1 (bases 1 to 69817)

Rounsley, S.D., Lin, X., Ketchum, K.A., Crosby, M.L., Brandon, R.C.,

Sykes, S.M., Kaul, S., Mason, T.M., Kerlavage, A.R., Adams, M.D.,

et al.

REFERENCE

1 (bases 1 to 69817)

Rounsley, S.D., Lin, X., Ketchum, K.A., Crosby, M.L., Brandon, R.C.,

Sykes, S.M., Kaul, S., Mason, T.M., Kerlavage, A.R., Adams, M.D.,

et al.

TITLE

Arabidopsis thaliana chromosome II BAC T29F13 genomic sequence

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 69817)

AUTHORS

Rounsley, S.D. and Lin, X.

TITLE

Direct Submission

JOURNAL

Submitted (07-NOV-1997) The Institute for Genomic Research, 9712

REFERENCE

Medical Center Dr., Rockville, MD 20850, USA, rounsley@tigr.org

AUTHORS

3 (bases 1 to 69817)

TITLE

Direct Submission

JOURNAL

Submitted (13-MAY-1998) The Institute for Genomic Research, 9712

REFERENCE

Medical Center Dr., Rockville, MD 20850, USA

AUTHORS

On May 13, 1998 this sequence version replaced gi:3126888.

TITLE

Address all correspondence to:

JOURNAL

Steve Rounsley

REFERENCE

The Institute for Genomic Research

AUTHORS

9712 Medical Center Dr.,

TITLE

Rockville, MD 20850,

JOURNAL

USA

REFERENCE

e-mail: rounsley@tigr.org

AUTHORS

BAC clone T29F13 is from Arabidopsis chromosome II and is near the

TITLE

molecular marker ve016.

JOURNAL

The orientation of the sequence is from SP6 to T7 end of the BAC

REFERENCE

clone.

AUTHORS

Genes were identified by a combination of three methods: Gene

TITLE

prediction programs including GRAIL (available by anonymous ftp

JOURNAL

from arthur.epm.ornl.gov), Genefinder (Phil Green, University of

REFERENCE

Washington), Genscan (Chris Burge,

AUTHORS

http://gnomic.stanford.edu/~chris/GENSCANW.html), and NetPlantGene

TITLE

(http://www.cbs.dtu.dk/netpene/cbsnetpene.html), searches of the

JOURNAL

complete sequence against a peptide database and the Arabidopsis

REFERENCE

EST database at TIGR (http://www.tigr.org/tdb/at/at.html).

AUTHORS

Annotated genes are named to indicate the level of evidence for

TITLE

their annotation. Genes with similarity to other proteins are named

JOURNAL

after the database hits. Genes without significant peptide

REFERENCE

similarity but with EST similarity are named as 'unknown' proteins.

AUTHORS

Genes without protein or EST similarity are named as 'unknown' proteins.

TITLE

than two gene prediction programs over most of their length are

JOURNAL

annotated as 'hypothetical' proteins. Genes encoding tRNAs are

REFERENCE

predicted by tRNAscan-SE (Sean Eddy,

AUTHORS

http://genome.wustl.edu/eddy/tRNAscan-SE/).

TITLE

identified by repeatmasker (Arian Smit,

JOURNAL

http://ftp.genome.washington.edu/RM/RepeatMasker.html). Regions of

REFERENCE

genomic sequence that are not annotated as genes but have predicted

AUTHORS

exons by GRAIL are annotated as misc features.

TITLE

Location/Qualifiers

JOURNAL

1. .69817

REFERENCE

/organism="Arabidopsis thaliana"

AUTHORS

/cultivar="Columbia"

TITLE

/db_xref="taxon:3702"

JOURNAL

/chromosome="II"

REFERENCE

/map="ve016"

AUTHORS

/clone="T29F13"

TITLE

complement(join(<1..273,386..516,635..1947,2188..3330))

JOURNAL

/gene="T29F13.1"

REFERENCE

complement(1..1207)

AUTHORS

/gene="T29F13.1"

TITLE

/note="overlap with BAC clone F1913 (AC004238:1..1207)"

JOURNAL

complement(<1..3330)

REFERENCE

/gene="T29F13.1"

AUTHORS

/note="F1913.1"

TITLE

complement(join(<1..273,386..516,635..1947,2188..3330))

JOURNAL

/gene="T29F13.1"

REFERENCE

/note="unknown protein, 3' partial"

AUTHORS

/codon_start=1

TITLE

/protein_id="AAC16269.1"

JOURNAL

/db_xref="PID:g3132480"

REFERENCE

/db_xref="GI:3132480"

AUTHORS

/translation="MATDGTGKNGSTVKSLEKEISLKFVLSLOOKLERNLKEKS

TITLE

ETKLLQDOASGREKEINELRLADSSFEEREHAFKELNKAALIVKDEEI

JOURNAL

EDIDPEVREISLVKNLASERQKTESERKKAESEKKADKYLSELYRNSAHTSS

REFERENCE

DLTLTNSLETVKQLEKQKTLKRAKRADESAKADQMKAEDVSKKFEIVRARN

AUTHORS

JOURNAL	Submitted (11-JUL-1998) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA	gene	complement(6205..9214)
REFERENCE	3 (bases 1 to 71097)	CDS	/gene="F15K9.3"
AUTHORS	Theologis,A.		complement(join(6205..6675,7214..9214))
JOURNAL	Direct Submission		/gene="F15K9.3"
REFERENCE	Submitted (24-OCT-1998) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA		/codon_start=1
AUTHORS	4 (bases 1 to 71097)		/evidence=not_experimental
JOURNAL	Theologis.		/product="F15K9.3"
	Direct Submission		/protein_id="AAC72106.1"
COMMENT	Submitted (07-NOV-1998) Plant Gene Expression Center, 800 Buchanan St., Albany, CA 94710, USA		/db_xref="PID:g3850566"
	On Oct 24, 1998 this sequence version replaced gi:3406036.		/db_xref="GI:3850566"
	This sequence is of BAC F15K9 from Arabidopsis thaliana chromosome 1. The sequence does not represent the sequence of the entire		/translation="MSIVQKEEMNGGLNVKVEAFVTSPOEKGKRNKRKLADPSQPNASSLTPEPVELPSPKPNHLSGNSGVEYSNQLQVEVSEVWDDPDPFACHLELLS
	insert of this clone. It is shorter by 26,439 bp because we submit		SNLTLFLDMKQIDIGYTDDEVILKAVSRCLRCCGGNNLLSNVITLSALKTIDEG
	only the unique sequence of the clone. In order to facilitate the		AGSDGYFEDQLQVLYSTLVMISLKEVRPSLSTVEAMWLLMCDLNLVQAFA."3D
	joining of overlapping clones in the future for creation of larger		GLVSSKLSDESLEGAENPNPKSDPDNPQSDPQPNRNNEPLKFGFNPTPNSEKKT
	contigs, we provide a small overlap (200 bp) between overlapping		QSSGTTTCGKEVCSGTVSCSGSTFTLVSDKLVSCRGKRTKKEIAMLKQKSCVEK
	submitted clones. The 5' end of this sequence overlaps by 200 bp		IRTSYSGSGVKAASFVSGFLLEKRVKSSSEFVPRNSSKITAIGVYKLSAEDSGC
	the 3' end of the sequence of BAC F21B7.		FVRNSKLDSPVVVDAKGYITALPARSVKSAKKKTCGSESVTLIPSAEKKSSSIP
FEATURES	Location/Qualifiers		STSEKKSSESEKASVAKLAPDIYAGIPYDAALGIYVPRDKDELILKLVPRVNDL
source	1..71097		QNELOQVTDWANQVKVEATGRLLKQDPKELKARKEREAEQYKKEQLLENTKRRLS
	/organism="Arabidopsis thaliana"		EMDFALKNATSOLEAKHTAHLRELSILKKEMAAKIKAVESAESFEAKERGRS
	/cultivar="Columbia"		LKDIHSEWGQIMLEELKGOREKVTYLOKEVTKAKNRQNOIEAALKOERTAKGLSA
	/db_xref="taxon:3702"		QASLIRKETLEALGKVEERIKGAEITDVKYVIDNIKRLEREISELKSDYSRII
	/chromosome="1"		ALKKSSSEKATRESLGMKPKVRERECVMCLSEMSVIFLPCAHOVLCFKCNQLHEK
	/clone="F15K9"		EGMWDPCSGRTIHRIRQARFARSG"
	/gene="F15K9.1"		complement(10217..11743)
	complement(59..241)		/gene="F15K9.4"
	complement(59..241)		complement(join(10217..10651,10741..10809,10909..10974,
	/gene="F15K9.1"		11054..11143,11249..11294,11404..11505,11583..11743))
	/note="This gene is cut off; it is continued from BAC		/gene="F15K9.4"
	F21B7."		/note="Similar to hypothetical protein SPAC2f7.14c
	/codon_start=1		g11052797 from Schizosaccharomyces pombe cosmid
	/evidence=not_experimental		g1250142."
	/protein_id="AAC72107.1"		/codon_start=1
	/db_xref="PID:g3850567"		/evidence=not_experimental
	/db_xref="GI:3850567"		/protein_id="AAC72108.1"
	/translation="MMVRKLQPLSQTKQVFERAIERLQSLSSANSASDVIVTDSI		/db_xref="PID:g3850568"
	VAAISAAEDDKDQVP"		/db_xref="GI:3850568"
	complement(1400..5428)		/translation="MMVRKLQPLSQTKQVFERAIERLQSLSSANSASDVIVTDSI
	/gene="F15K9.2"		PVNHDDAFVLGHGTSEVDGELLATVCGVVERVDKLVVVRTLRARYKPVGDIVGVRVI
	complement(join(1400..1882,1977..2142,2440..2702,		EVAQKRWRFNLFNQDGLVMSNMMPDGIQRRTSVDELNWRNIFVEHVDVYCAVRN
	2826..2942,3029..3312,3461..3643,3756..4976,5083..5428))		FQHDGSLQARSAQKYKLEKQQLKVDPIYLRKSKHFFHYVESGLDIIICNGFIW
	/gene="F15K9.2"		VGEHVRDPMADIDDKDEEMISSSTGKEQSHIPTLETRQICRIGNAIRVLSNLGFT
	/note="Contains similarity to gb AB011110 KIAA0538 protein		12150..13562
	from Homo sapiens brain and to phospholipid-binding domain		/gene="F15K9.5"
	C2 PF100168. ESTs gb AA585988 and gb T04384 come from		12150..13562
	this gene."		/gene="F15K9.5"
	/codon_start=1		/note="ESTs gb T21276, gb T45403, and gb AA586113 come
	/evidence=not_experimental		from this gene."
	/protein_id="AAC72128.1"		/codon_start=1
	/db_xref="PID:g3850588"		/evidence=not_experimental
	/db_xref="GI:3850588"		/protein_id="AAC72109.1"
	/translation="MKIQVRVVEARNLPAMDNLGCFSDPYVRLQLGKORSRTKVKKNL		/db_xref="PID:g3850569"
	NPKTEFSGVDLDELVSVLDEKYNFNDRFVQVRVSVSLVFAENQSLGTVMY		/db_xref="GI:3850569"
	PLNPKKGSKDCDEILLKCYFSQKNSVLDITSSGDDTSASRSPDLRLESPIDSTCA		/translation="MNFVKSVFTEDLPDPETESDSPKHEEHEPEQHEPEQSESN
	SPSRSDASSIPQTFAGRTQIFQKNAITATPQSSRSIDASDLSEISKPVSLEL		DDGWSFGGLMKTLATRESVIETRRDLEFGTGLKETEVAOGSLGTVCGAIDELG
	SEDSSTSEELLKAMESKDGPSPNSLGGVVPQFLWISPDNLNIVLFASDSDF		NTVLKGAEIIAIOGKEAILAAGNESDDNNSSQSGFRROSFSSKPYSRFDAQIRAVQ
	YASLTELQGTVEQVIGPWKAENDGESVRYVYLKAATKLIKAVKTEQOTLYKADGE		GDLTNCEEPEDSDYKKESASLDGKAEMKELLENGDMKGVYKRVFPMVDHET
	YAVLASVATPDVPFGGTFKVEVLYCISPGPELSPGQCSRLVYVSWRLNFIQSTMRG		FWRFYFRVKNLQKQEDLRANLVKRAISLDDEEELSWDIDDEESSEKVVATKDVSR
	LIENGARQGLKDNFEQANLLAQSVKPVDSKDIGNKEALSSLOQAPQSDMKLAVQI		LKLEGNMGDGVDSVETKDEVESTYSVAKVTODEVTSADSVTEVSNVGLTKDQSE
	FANTVLSTFLIGIYFVHVIIPAISAIQGLEFNLGLDPSIGFVYSGDVLVQCERY		EKETDSEEEYPEEKSFDAAPPASDEAPIQDSVKPTDEAPIQDSVKPKSDEAAPSD
	LOLSIRPMQARKSGDHGIKAGHGMLLTVALLIEVDLAADVPDGHCDPIYVTSNG		SAPDVAASSTQQPSSEEDLGWDEIDEMSSIDGKETSRSGSGSPNRAELKRLSAAED
	KRTSSIKFQKSNPQWNEIEFFDAMADPPSVLNVVEFDGDFDPEASLGHAEVNFVR		EDLSWDIDDEDEESSSKA"
	SNISDLADVWPLQKLAQACQSLHLRIFLDHTGGGVDVDRDYLNMKEKEVGKKINVR		complement(15169..16684)
	SPQNSAFQKLPQPEFLINDFTCHLRKMPQLQGRLEFLSARIVGYASIFGNKTFE		/gene="F15K9.6"
	FFLWEDIEETQVLPPTLASGSPVIVMTLRPNRGLDRIAGKTHDEEGRKLFHPHSFV		complement(join(15169..15658,15730..15801,15895..16237,
	SNFYAQKTIKALWAKSLTPEQKQVVEEESQKLSQESGLFLGVDVDRFSEVFSIT		16319..16499,16613..16684))
	LPVPVFFMFLFPGKGVDRKMARAGQSCSPWSEKDDVYRDKRISRYR		/gene="F15K9.6"
	GEVTSQOKSLVLPKGNLWVEVMTLHGVPGLGDYFNLHRYQMESTSKPKTTVVRVY		/codon_start=1
	FGIEWLAKSTRQKRVTKNVLNQLDRLKMFTFGFLEKEYESSRQQQQVY"		/evidence=not_experimental
			/product="F15K9.6"
			/protein_id="AAC72110.1"

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 05:27:13 : Search time 213.04 Seconds
(without alignments)
23.488 Million cell updates/sec

Title: US-09-037-472-2
Perfect score: 20
Sequence: 1 GTACCTCCGAGTATACATT 20

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	20	100.0	20	V32390	Interleukin-1-alpha
2	14.4	72.0	1578	V62388	Human 7-transmembrane
3	14.4	72.0	1578	V71117	cDNA encoding a hu
4	14.2	71.0	13585	T11549	Tumour rejection a
5	14.2	71.0	2436	V28436	Aspergillus niger
6	14.2	71.0	110000	V21209_07	Continuation (8 of
7	14.2	71.0	278	V90269	EST clone D1160. N
8	14.2	71.0	6022	V74403	Staphylococcus aur
9	14.2	71.0	5069	X13259	Enterococcus faeca
10	14.2	71.0	5137	V12500	Mouse metastasis-a
11	13.8	69.0	125	V65134	Transforming growt
12	13.8	69.0	125	V65155	Transforming growt
13	13.8	69.0	125	V65156	Transforming growt
14	13.8	69.0	116	V65157	Transforming growt
15	13.8	69.0	123	V65158	Transforming growt
16	13.8	69.0	117	V65159	Transforming growt
17	13.8	69.0	123	V65160	Transforming growt
18	13.8	69.0	115	V65161	Transforming growt
19	13.8	69.0	116	V65162	Transforming growt
20	13.8	69.0	116	V65166	Transforming growt
21	13.8	69.0	78	V65169	Transforming growt
22	13.8	69.0	5621	X13508	Enterococcus faeca
23	13.6	68.0	897	V31988	Guinea pig MBP-1 g
24	13.6	68.0	1720	V78916	Non-typable Haemop
25	13.6	68.0	3756	V99863	Mouse LTBP-3 gene.
26	13.6	68.0	1193	V47220	Desulfurococcus st
27	13.6	68.0	7115	V43042	Mus musculus Cgamm
28	13.6	68.0	289	V68924	DNA molecule encod
29	13.6	68.0	2241	V68233	Nucleotide sequenc
30	13.6	68.0	13715	V74506	Staphylococcus aur
31	13.6	68.0	2311	X13205	Enterococcus faeca
32	13.6	68.0	8033	X13058	Enterococcus faeca
33	13.6	68.0	1076	X13772	Enterococcus faeca
34	13.6	68.0	1560	X00225	Clostridium chauvo
35	13.4	67.0	992	V04091	Clone FT36 contain
36	13.2	66.0	3685	V81712	Clone contg. hom-t
37	13.2	66.0	877	V81774	DNA coding sequenc
38	13.2	66.0	12492	V70543	Complete sequence
39	13.2	66.0	4839	V06827	Alpha amylase pull
40	13.2	66.0	997	V70951	Sequence of enhanc
41	13.2	66.0	2259	V01218	Sequence encoding
42	13.2	66.0	2615	V60971	Sequence encoding
43	13.2	66.0	2350	V021184	Encodes T lymphocy

C 44 13.2 66.0 1914 1 Q22500 Encodes T lymphocy
C 45 13.2 66.0 2350 1 V63460 Human Leu8 antigen

ALIGNMENTS

RESULT 1

V32390 standard; DNA; 20 BP.
AC V32390;
DT 11-SEP-1998 (first entry)
DE Interleukin-1-alpha primer 2 (-888/-869).
KW IL-1-alpha; genetic polymorphism; PCR; primer; amplification;
KW sight threatening diabetic retinopathy; interleukin-1-alpha;
KW Interleukin-1-beta; Interleukin-1RN; ss.
OS Synthetic.
OS Homo sapiens.
PN WO9815653-A1.
PD 16-APR-1998.
PF 08-OCT-1997; G02790.
PR 10-OCT-1996; GB-021129.
PA (DUFFE/) DUFF G.
PA (RENN/) RENNIE I.
PA (RICH/) RICHARDSON R.
PI Duff G, Rennie I, Richardson R;
DR WPI: 98-240835/21.
PT Predicting increased risk of sight-threatening diabetic retinopathy
PT - comprises identifying genetic polymorphism pattern for genes
PT IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical
PT symptoms occur
CC Claim 2; Page 33; 41pp; English.
CC Interleukin-1-alpha (IL-1-alpha) primers 2 and 1 (V32389) were used to
CC amplify the IL-1-alpha gene region to identify single base variation
CC polymorphism of C/T at base 889. The invention claims to provide a
CC method for predicting the risk of sight threatening diabetic retinopathy.
CC The method involves isolating DNA from a patient and determining the DNA
CC polymorphism pattern of the genes that code for interleukin-1-alpha,
CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern
CC identified is then compared with controls of known DNA polymorphism
CC patterns thereby identifying patients carrying a genetic polymorphism
CC associated with increased risk of sight threatening diabetic retinopathy.
CC The method may be able to identify diabetic patients at risk before the
CC clinically detectable disorders occur. Polymorphism pattern
CC determination of IL genes involved PCR reactions using primers V32389-
CC V32398. The method is also claimed to be useful in conjunction with
CC identification of other genes associated with sight threatening diabetic
CC retinopathy in genomic DNA and therefore, in identifying diabetic
CC patients expressing multiple risk patterns.
SQ Sequence 20 BP: 5 A; 5 C; 3 G; 7 T;

Query Match 100.0%; Score 20; DB 1; Length 20;
Best Local Similarity 100.0%; Pred. No. 0.025;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 GTACCTCCGAGTATACATT 20

Db 1 GTACCTCCGAGTATACATT 20

RESULT 2

V62388 standard; CDNA; 1578 BP.

AC V62388;
DT 19-JAN-1999 (first entry)
DE Human 7-transmembrane receptor polypeptide, HMTMF81 encoding CDNA.
KW HMTMF81; 7-transmembrane receptor; treatment; infection; bacteria; pain;
KW fungal; protozoan; viral; human immune deficiency virus; HIV-1; HIV-2;
KW cancer; anorexia; bulimia; asthma; Parkinson's disease; heart failure;
KW hypotension; hypertension; urinary retention; osteoporosis; allergy;
KW angina pectoris; myocardial infarction; asthma; allergy; ulcer; anxiety;
KW prostatic hypertrophy; psychotic disorder; neurological disorder; human;

KW schizophrenia; manic depression; delirium; dementia; mental retardation;
OS dyskinesias; Huntington's disease; Gilles de la Tourette's syndrome; ss.
FH Homo sapiens.
FT Key Location/Qualifiers
FT CDS 467..1480
FT FT /*tag= a
FT FT /product= "HMTMF81 polypeptide"
PN EP-874047-A2.
PD 28-OCT-1998.
PF 20-APR-1998; 303008.
PR 19-MAR-1998; US-844795.
PR 22-APR-1997; US-844795.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.
PI Ames R, Chambers J, Ellis C, Foley J, Halsey W,
PI Sarau H, Sathe G;
DR WPI: 98-544641/47.
DR P-PSDB: W75799.
PT DNA encoding 7-trans-membrane receptor polypeptide HMTMF81 - useful
PT in treatment of e.g. infections such as bacterial, fungal, protozoan
PT and viral infections, particularly HIV, cancers and bulimia etc.
PS Claim 5; Pages 7-8; 22pp; English.
CC This cDNA encodes a human 7-transmembrane receptor polypeptide, HMTMF81.
CC The HMTMF81 polypeptides and polynucleotides can be used in the treatment
CC of infections such as bacterial, fungal, protozoan and viral infections.
CC They can be used particularly for treatment of infections caused by
CC (human immune deficiency virus) HIV-1 or HIV-2, pain, cancers, anorexia,
CC bulimia, asthma, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC myocardial infarction, ulcers, asthma, allergies, benign prostatic
CC hypertrophy, and psychotic and neurological disorders, including
CC anxiety, schizophrenia, manic depression, delirium, dementia, severe
CC mental retardation and dyskinesias such as Huntington's disease or
CC Gilles de la Tourette's syndrome.
CC Sequence 1578 BP; 480 A; 322 C; 305 G; 471 T;
SQ

Query Match 72.0%; Score 14.4; DB 1; Length 1578;
Best Local Similarity 93.8%; Pred. No. 48;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCGAGTATACAT 19
||||| |||||||
Db 636 CCTTCCAGTATACAT 651

RESULT 3
V71117
ID V71117 standard; cDNA: 1578 BP.
AC V71117;
DE cDNA encoding a human 7-transmembrane receptor designated HMTMF81.
KW Human; 7-transmembrane receptor protein; HMTMF81; infection;
KW HIV; pain; cancer; anorexia; bulimia; asthma; Parkinson's disease;
KW acute heart failure; hypotension; hypertension; urinary retention;
KW osteoporosis; angina pectoris; myocardial infarction; ulcer; asthma;
KW benign prostatic hypertrophy; neurological disorder; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 467..1480
FT FT /*tag= a
FT FT /product= HMTMF81
PN EP-878542-A2.
PD 18-NOV-1998.
PF 27-OCT-1997; 308560.
PR 22-APR-1997; US-844795.
PA (SMIK) SMITHKLINE BEECHAM CORP.
PI Ellis CE, Halsey WS, Sathe GW;
DR WPI: 98-585747/50.
DR P-PSDB: W85047.
PT DNA encoding 7-transmembrane receptor polypeptide HMTMF81 - useful
PT for treatment of, e.g. HIV infections, pain, cancers, myocardial
PT infarction and acute heart failure

PS Claim 3; Pages 17-18; 20pp; English.
CC The present sequence encodes a human 7-transmembrane receptor protein
CC designated HMTMF81. HMTMF81 polypeptides and polynucleotides can be
CC used in the treatment of infections such as bacterial, fungal,
CC protozoan and viral infections, particularly infections caused by
CC HIV-1 or HIV-2. They can also be used to treat pain, cancers, anorexia,
CC bulimia, asthma, Parkinson's disease, acute heart failure, hypotension,
CC hypertension, urinary retention, osteoporosis, angina pectoris,
CC myocardial infarction, ulcers, asthma, allergies, benign prostatic
CC hypertrophy and psychotic and neurological disorders, including anxiety,
CC schizophrenia, manic depression, delirium, dementia, severe mental
CC retardation and dyskinesias such as Huntington's disease or Gilles
CC de la Tourette's syndrome.
CC Sequence 1578 BP; 480 A; 322 C; 305 G; 471 T;
SQ

Query Match 72.0%; Score 14.4; DB 1; Length 1578;
Best Local Similarity 93.8%; Pred. No. 48;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 4 CCTTCGAGTATACAT 19
||||| |||||||
Db 636 CCTTCCAGTATACAT 651

RESULT 4
T11549
ID T11549 standard; DNA: 13585 BP.
AC T11549;
DE Tumour rejection antigen precursor coding sequence.
DE Tumour rejection antigen precursor; TRAP; TRA; melanoma; cancer;
KW Tumour rejection antigen precursor; TRAP; TRA; melanoma; cancer;
KW tumour; treatment; detection; vaccine; HLA-A2; adoptive transfer;
KW T cell; T lymphocyte; human leukocyte antigen; ds.
OS Homo sapiens.
FH Key Location/Qualifiers
FT misc_feature 2685
FT /*tag= a
FT /note= "Unidentified nucleotide."
FT misc_feature 9422..9456
FT /*tag= b
FT /note= "This region is 4.7-5.3 kilobases in length
FT but the sequence of this region has not been
FT deduced."
FT misc_feature 11539
FT /*tag= c
FT /note= "Unidentified nucleotide."
PN WO9601557-A1.
PD 25-JAN-1996.
PF 27-JUN-1995; U08153.
PR 08-JUL-1994; US-272351.
PR 10-JAN-1995; US-370319.
PA (LUDW-) LUDWIG INST CANCER RES.
PI Boon-fallier T, Brichard V, Coullie P, De PLAEN E;
PI Traversari C, Van PEL A, Wolfel T;
DR WPI: 96-097390/10.
PT Genomic DNA encoding a tumour rejection antigen precursor -
PT processed to antigen presented by HLA-A2, useful for treating or
PT diagnosing melanoma
PT Claim 3; Page 18-22; 41pp; English.
PS The tumour rejection antigen precursor (TRAP) is processed to a
CC tumour rejection antigen (TRA) presented by HLA-A2 molecules. TRA
CC is used to generate cytotoxic T lymphocytes for treating cancer (esp.
CC melanoma). It can also be used to raise specific antibodies, and
CC when complexed with HLA-A2, it can be used to produce vaccines.
CC Cytotoxic T lymphocytes so generated can be used in adoptive
CC transfer or generated, or they can be generated in vivo by using a
CC vector containing the appropriate gene or using TRA or TRAP together
CC with an adjuvant that facilitates gene into HLA-A2 presenting cells.
CC Diagnostic methods involving the detection of expression of TRAP
CC can be used in the detection of cancers.
CC Sequence 13585 BP; 3827 A; 2859 C; 2968 G; 3894 T;
SQ

Query Match 71.0%; Score 14.2; DB 1; Length 13585;
 Best Local Similarity 84.2%; Pred. No. 78;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TACCTTCGAGTATACATT 20
 ||||| ||||| ||||| |||||

Db 10375 TACCTTCGAGTATGATT 10393

RESULT 5
 V28496

ID V28496 standard; DNA; 2436 BP.

AC V28496;
 DT 28-APR-1998 (first entry)
 DE Aspergillus niger ferulic acid esterase gene.
 KW Ferulic acid esterase; esterolytic enzyme; feed supplement;
 KW fabric; yarn; textile; ss.
 OS Aspergillus niger.

FH Key Location/Qualifiers

FT Key 519..1421

FT CDS /*tag= a

FT /*note= "Includes an intron"

FT sig_peptide 519..581

FT /*tag= b

FT intron 980..1037

FT /*tag= c

PN WO9814594-A2.

PD 03-APR-1998.

PF 29-SEP-1997; U17614.

PR 30-SEP-1996; US-722713.

PA (GEMV) GENENCOR INT INC.

PI Borneman WS, Bower BS;

DR WPI: 98-240093/21.

DR P-PSDB; W56728.

PT New esterase from Aspergillus niger - useful for, e.g. feed and food
 PT processing, treating fabrics or for producing antioxidant,
 PT photoprotective or anti-inflammatory agents

PS Claim 6: Fig 5A-C: 40pp; English.

CC This DNA sequence corresponds to the gene encoding a novel 39 kDa
 CC ferulic acid esterase (FAE, see W56728) of Aspergillus niger. It
 CC was isolated from A. niger genomic DNA using a partial gene
 CC fragment (see V28496) as probe. The invention provides vectors
 CC that include the DNA sequence, host cells transformed with the DNA
 CC or vectors, fermentation broths comprising such host cells, and

CC esterase proteins expressed by the host cells. The FAE can cleave
 CC the ester linkage of phenolic esters. It can be used in a claimed
 CC feed supplement and in a claimed process of treating fabric yarn or
 CC textiles. FAE gene fragments can be used as probes to isolate
 CC homologous sequences having esterase activity.

SQ Sequence 2436 BP; 633 A; 615 C; 574 G; 614 T;

Query Match 71.0%; Score 14.2; DB 1; Length 2436;
 Best Local Similarity 84.2%; Pred. No. 65;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTACCTTCGAGTATACAT 19
 || ||| ||||| ||||| ||

Db 300 GTGCTCCGAGTATACATT 318

RESULT 6

V21209_07

Continuation (8 of 17) of V21209 from base 700001 (Methanococcus jannaschii circular chr
 WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

WP Fragment Name Begin End

WP V21209_00 1 110000

WP V21209_01 100001 210000

WP V21209_02 200001 310000

WP V21209_03 300001 410000

WP V21209_04 400001 510000

WP V21209_05 500001 610000

WP V21209_06 600001 710000
 WP V21209_07 700001 810000
 WP V21209_08 800001 910000
 WP V21209_09 900001 1010000
 WP V21209_10 1000001 1110000
 WP V21209_11 1100001 1210000
 WP V21209_12 1200001 1310000
 WP V21209_13 1300001 1410000
 WP V21209_14 1400001 1510000
 WP V21209_15 1500001 1610000
 WP V21209_16 1600001 1664976

Query Match 71.0%; Score 14.2; DB 1; Length 110000;
 Best Local Similarity 84.2%; Pred. No. 96;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTACCTTCGAGTATACAT 19
 || ||||| ||||| |||||

Db 44008 GGAACCTCCGAGATACAT 44026

RESULT 7

V90269

ID V90269 standard; CDNA; 278 BP.

AC V90269;

DT 15-FEB-1999 (first entry)

DE EST clone D1160.

KW Human; secreted protein; expressed sequence tag; EST; haematopoiesis;

KW tissue growth; activin; inhibin; chemotaxis; chemokinesis; haemostatic;

KW receptor; ligand; thrombolytic; anti-inflammatory; cadherin; anti-tumour;

KW gene therapy; ss.

OS Homo sapiens.

PN WO9845436-A2.

PD 15-OCT-1998.

PF 10-APR-1998; U06955.

PR 10-APR-1997; US-838821.

PA (GEMV) GENETICS INST INC.

PI Racie LA, Spaulding V, Treacy M;

DR WPI: 99-070077/06.

PT New polynucleotides encoding human secreted proteins - derived from
 PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
 PT ovary, pituitary, retina and colon cDNA libraries.

PS Claim 1; Page 486; 618pp; English.

CC The present sequence represents a human expressed sequence tag (EST).

CC The polynucleotide, which is a secreted EST, and the encoded protein

CC are predicted to have useful biological activities which would make

CC them suitable for treating, preventing or ameliorating medical

CC conditions in humans and animals, although no supporting data is

CC given. Suggested activities include nutritional activity, immune

CC stimulating or suppressing activity, haematopoiesis regulating

CC activity, tissue growth activity, activin/inhibin activity,

CC chemotactic/chemokinetic activity, haemostatic and thrombolytic

CC activity, receptor/ligand activity, anti-inflammatory activity,

CC cadherin/tumour invasion suppressor activity, tumour inhibition

CC activity. The polynucleotide may also be useful for gene therapy.

SQ Sequence 278 BP; 90 A; 71 C; 46 G; 71 T;

Query Match 71.0%; Score 14.2; DB 1; Length 278;
 Best Local Similarity 84.2%; Pred. No. 51;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 TACCTTCGAGTATACATT 20
 || ||| || ||||| |||||

Db 250 TGCTACCCAGTATACATT 268

RESULT 8

V74403

ID V74403 standard; DNA; 6022 BP.

AC V74403;

DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #92.
 KW Computer readable medium; vaccine; S aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome; ds.
 OS Staphylococcus aureus.
 FH Key Location/Qualifiers
 FT misc_feature 421..480
 FT /*tag= a
 FT /*note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc_feature 2221..2280
 FT /*tag= b
 FT /*note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc_feature 4021..4080
 FT /*tag= c
 FT /*note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT misc_feature 5821..5880
 FT /*tag= d
 FT /*note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence."
 FT EP-786519-A2.
 PN 30-JUL-1997.
 PD 07-JAN-1997; 100117.
 PF 05-JAN-1996; US-009861.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 PI WPI; 97-374922/35.
 DR Polynucleotide(s) and proteins derived from Staphylococcus aureus -
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 PS Claim 1: Page 569-573: 3271pp; English.
 CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 6022 BP; 2021 A; 993 C; 805 G; 1960 T;

Query Match 71.0%; Score 14.2; DB 1; Length 6022;
 Best Local Similarity 84.2%; Pred. No. 71;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TACCTTCGAGTATACATT 20
 Db 5507 TACTTCTCTGTATACATT 5525

RESULT 9
 X13259/c
 ID X13259 standard; DNA; 5069 BP.
 AC X13259;
 DT 19-MAR-1999 (first entry)
 DE Enterococcus faecalis genome contig SEQ ID NO:322.
 KW Enterococcus faecalis; contig; detection; Enterococcal infection;
 KW vaccine; attenuation; computer readable medium; ds.
 OS Enterococcus faecalis.
 PN WO980555-A2.
 PD 12-NOV-1998.
 PF 04-MAY-1998; U08985.
 PR 14-NOV-1997; US-066009.
 PR 06-MAY-1997; US-044031.
 PR 16-MAY-1997; US-046655.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Dillon PJ, Kunsch CA;
 PI WPI; 99-045171/04.
 PT New isolated Enterococcus faecalis polynucleotides and polypeptides
 PT - used to develop products for the detection of Enterococcus and for
 PT use in vaccines for prevention or attenuation of Enterococcus
 PT infection.
 PS Claim 1: Page 1412-1415; 2084pp; English.
 CC A computer readable medium has been developed which has recorded on it
 CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
 CC X12938 to X13919 represent these nucleotide sequences which are primary
 CC nucleotide sequences, also known as contigs. The computer-based system
 CC can identify fragments of the Enterococcus faecalis genome with
 CC commercial importance. The products can be used to detect the presence
 CC of Enterococcus faecalis in samples. They can also be used for
 CC diagnosing Enterococcal infection in an animal and monitoring
 CC progression of disease, and for identifying agents which can be used to
 CC modulate the growth or pathogenicity of Enterococcus faecalis, or
 CC another related organism, in vivo or in vitro. In particular the
 CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
 CC can be used in vaccines to prevent or attenuate an Enterococcal
 CC infection.
 SQ Sequence 5069 BP; 1620 A; 892 C; 1171 G; 1384 T;
 Query Match 71.0%; Score 14.2; DB 1; Length 5069;
 Best Local Similarity 84.2%; Pred. No. 70;
 Matches 16; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
 QY 2 TACCTTCGAGTATACATT 20
 Db 1118 TCCCTCCGAGTATACATT 1100
 RESULT 10
 V12500/c
 ID V12500 standard; cDNA; 5137 BP.
 AC V12500;
 DT 27-APR-1998 (first entry)
 DE Mouse metastasis-associated antigen p150 cDNA.
 KW Antigen; p150; metastasis; tumour; cancer; diagnosis; therapy;
 KW mouse; ss.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT CDS 128..4159
 FT /*tag= a
 FT WO9741221-A1.
 PN 06-NOV-1997.
 PD 18-APR-1997; E01963.
 PR 28-FEB-1997; GB-004161.
 PR 29-APR-1996; US-016487.
 PA (NOVS) NOVARTIS AG.
 PI Bachmann F;
 PI WPI; 97-549725/50.
 DR P-PSDB; W31866.
 PT Mouse and human metastasis-associated p150-derived antigens - useful
 PT for raising antibodies for therapy and diagnosis of tumours

PS Claim 4; Page 41-51; 11pp; English.
 CC This cDNA clone codes for novel murine 150 kDa cell surface
 CC metastasis-associated antigen (see W31866), designated p150. The
 CC p150 protein was identified by raising antisera to membrane
 CC associated proteins and assessing cross-reactivity with transformed
 CC host cells. An oligonucleotide based on an isolated peptide of
 CC p150 was then used to screen a mouse melanoma Bi6F1 library,
 CC yielding a partial p150 cDNA clone. The missing 5' sequence was
 CC obtained by RACE. The cDNA sequence (see T93627) of the human
 CC homologue of p150 (see W31867) was also obtained. A claimed method
 CC for determining the metastatic potential of a tumour cell comprises
 CC assessing the level of p150 expression in the cell, a higher level
 CC of p150 expression being indicative of a lower state of
 CC differentiation and consequently a higher metastatic potential.
 CC p150 nucleic acids can be used in the recombinant production of
 CC p150 polypeptides, as well as in methods for the diagnosis and
 CC gene therapy of tumours
 SQ Sequence 5137 BP; 1664 A; 924 C; 1355 G; 1194 T;

Query Match 70.0%; Score 14; DB 1; Length 5137;
 Best Local Similarity 100.0%; Pred. No. 90;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 CCTTCCGAGTATAC 17
 Db 701 CCTTCCGAGTATAC 688
 |||||

RESULT 11
 T65154
 ID T65154 standard; RNA; 125 BP.
 AC T65154;
 DT 10-SEP-1997 (first entry)
 DE Transforming growth factor beta binding ligand lib3 13.
 KW Transforming; growth factor; beta; TGF-beta; binding ligand;
 KW identification; SELEX; anti-mitogenic; inhibition; cell;
 KW Systematic Evolution of Ligands by Exponential enrichment;
 KW epithelial; proliferation; diagnosis; treatment; fibroids;
 KW kidney; lung; liver; dermal scarring; restenosis; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 1..125
 FT /*tag= a
 FT /note= "all pyrimidines are 2'-fluoro modified"

PN WO9638579-A1.
 PD 05-DEC-1996.
 PF 30-MAY-1996; U08014.
 PR 02-JUN-1995; US-458424.
 PR 02-JUN-1995; US-458423.
 PR 05-JUN-1995; US-465594.
 PR 05-JUN-1995; US-465591.
 PR 07-JUN-1995; US-479783.
 PR 07-JUN-1995; US-479725.
 PR 20-MAR-1996; US-618693.
 PA (NEXS-) NEXSTAR PHARM INC.
 PI Gold L, Janjic N, Pagratis N, Ringquist S, Toothman PJ;
 DR WPI; 97-034387/03.
 PT Identification of nucleic acid ligands to TGF-beta, PDGF and hKGF -
 PT using SELEX, used in the diagnosis and treatment of proliferative
 PT disorders
 PS Claim 15; Page 108; 209pp; English.
 CC The present sequence, a transforming growth factor beta (TGF-beta)
 CC binding ligand, was identified by Systematic Evolution of Ligands
 CC by Exponential enrichment (SELEX). Briefly a candidate mixture of
 CC nucleic acids was contacted with TGF-beta, and nucleic acids
 CC having an increased affinity to TGF-beta partitioned from the
 CC remainder of the mixture. The partitioned nucleic acids were then
 CC amplified to yield a mixture of nucleic acids enriched for
 CC sequences with higher affinity and specificity for binding to
 CC TGF-beta. The ligand is anti-mitogenic and may be used to inhibit
 CC epithelial cell proliferation, or in the diagnosis and treatment of
 CC TGF-beta mediated pathological conditions, e.g. fibrotic conditions
 CC such as fibroids of the kidney, lung and liver and more acute
 CC conditions such as dermal scarring and restenosis.
 SQ Sequence 125 BP; 25 A; 28 C; 46 G; 26 U;

Query Match 69.08%; Score 13.8; DB 1; Length 125;
 Best Local Similarity 64.7%; Pred. No. 77;
 Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCTTCGAGTATACA 18
 Db 36 UAGCUUCCGAGUAGACA 52
 :|:|||||:|

RESULT 13

CC such as fibroids of the kidney, lung and liver and more acute
 CC conditions such as dermal scarring and restenosis.
 SQ Sequence 125 BP; 25 A; 27 C; 46 G; 27 U;

Query Match 69.0%; Score 13.8; DB 1; Length 125;
 Best Local Similarity 64.7%; Pred. No. 77;
 Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCTTCGAGTATACA 18
 Db 36 UAGCUUCCGAGUAGACA 52
 :|:|||||:|

RESULT 12

T65155
 ID T65155 standard; RNA; 125 BP.
 AC T65155;
 DT 10-SEP-1997 (first entry)
 DE Transforming growth factor beta binding ligand lib3 3.
 KW Transforming; growth factor; beta; TGF-beta; binding ligand;
 KW identification; SELEX; anti-mitogenic; inhibition; cell;
 KW Systematic Evolution of Ligands by Exponential enrichment;
 KW epithelial; proliferation; diagnosis; treatment; fibroids;
 KW kidney; lung; liver; dermal scarring; restenosis; ss.
 OS Synthetic.
 FH Key Location/Qualifiers
 FT misc_feature 1..125
 FT /*tag= a
 FT /note= "all pyrimidines are 2'-fluoro modified"

PN WO9638579-A1.
 PD 05-DEC-1996.
 PF 30-MAY-1996; U08014.
 PR 02-JUN-1995; US-458424.
 PR 02-JUN-1995; US-458423.
 PR 05-JUN-1995; US-465594.
 PR 05-JUN-1995; US-465591.
 PR 07-JUN-1995; US-479783.
 PR 07-JUN-1995; US-479725.
 PR 20-MAR-1996; US-618693.
 PA (NEXS-) NEXSTAR PHARM INC.
 PI Gold L, Janjic N, Pagratis N, Ringquist S, Toothman PJ;
 DR WPI; 97-034387/03.
 PT Identification of nucleic acid ligands to TGF-beta, PDGF and hKGF -
 PT using SELEX, used in the diagnosis and treatment of proliferative
 PT disorders
 PS Claim 15; Page 108; 209pp; English.
 CC The present sequence, a transforming growth factor beta (TGF-beta)
 CC binding ligand, was identified by Systematic Evolution of Ligands
 CC by Exponential enrichment (SELEX). Briefly a candidate mixture of
 CC nucleic acids was contacted with TGF-beta, and nucleic acids
 CC having an increased affinity to TGF-beta partitioned from the
 CC remainder of the mixture. The partitioned nucleic acids were then
 CC amplified to yield a mixture of nucleic acids enriched for
 CC sequences with higher affinity and specificity for binding to
 CC TGF-beta. The ligand is anti-mitogenic and may be used to inhibit
 CC epithelial cell proliferation, or in the diagnosis and treatment of
 CC TGF-beta mediated pathological conditions, e.g. fibrotic conditions
 CC such as fibroids of the kidney, lung and liver and more acute
 CC conditions such as dermal scarring and restenosis.
 SQ Sequence 125 BP; 25 A; 28 C; 46 G; 26 U;

Query Match 69.08%; Score 13.8; DB 1; Length 125;
 Best Local Similarity 64.7%; Pred. No. 77;
 Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCTTCGAGTATACA 18
 Db 36 UAGCUUCCGAGUAGACA 52
 :|:|||||:|

```

T65156
ID T65156 standard; RNA; 125 BP.
AC T65156;
DE 10-SEP-1997 (first entry)
KW Transforming growth factor beta binding ligand lib3 4.
KW Transforming growth factor; beta; TGF-beta; binding ligand;
KW Identification; SELEX; anti-mitogenic; inhibition; cell;
KW Systematic Evolution of Ligands by Exponential enrichment;
KW epithelial; proliferation; diagnosis; treatment; fibroids;
KW kidney; lung; liver; dermal scarring; restenosis; ss.
OS Synthetic.
FH Key Location/Qualifiers
FT misc_feature 1..125
FT FT /*tag= a
FT FT /note= "all pyrimidines are 2'-fluoro modified"

PN WO9638579-A1.
PD 05-DEC-1996.
PF 30-MAY-1996; U08014.
PR 02-JUN-1995; US-458424.
PR 02-JUN-1995; US-458423.
PR 05-JUN-1995; US-465594.
PR 05-JUN-1995; US-465591.
PR 07-JUN-1995; US-479783.
PR 07-JUN-1995; US-479725.
PR 20-MAR-1996; US-618693.
PA (NEXS-) NEXSTAR PHARM INC.
PI Gold L, Janjic N, Pagratis N, Ringquist S, Toothman PJ;
DR WPI; 97-034387/03.
PT Identification of nucleic acid ligands to TGF-beta, PDGF and hKGF
PT using SELEX, used in the diagnosis and treatment of proliferative
PT disorders
PS Claim 15; Page 109; 209pp; English.
CC The present sequence, a transforming growth factor beta (TGF-beta)
CC binding ligand, was identified by Systematic Evolution of Ligands
CC by Exponential enrichment (SELEX). Briefly a candidate mixture of
CC nucleic acids was contacted with TGF-beta, and nucleic acids
CC having an increased affinity to TGF-beta partitioned from the
CC remainder of the mixture. The partitioned nucleic acids were then
CC amplified to yield a mixture of nucleic acids enriched for
CC sequences with higher affinity and specificity for binding to
CC TGF-beta. The ligand is anti-mitogenic and may be used to inhibit
CC epithelial cell proliferation, or in the diagnosis and treatment of
CC TGF-beta mediated pathological conditions, e.g. fibrotic conditions
CC such as fibroids of the kidney, lung and liver and more acute
CC conditions such as dermal scarring and restenosis.
SQ Sequence 116 BP; 24 A; 27 C; 44 G; 21 U;

Query Match 69.0%; Score 13.8; DB 1; Length 116;
Best Local Similarity 64.7%; Pred. No. 77;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCTTCCGAGTATACA 18
Db 35 UAGCUUCCGAGUAGACA 51

RESULT 15
T65158
ID T65158 standard; RNA; 123 BP.
AC T65158;
DE 10-SEP-1997 (first entry)
DE Transforming growth factor beta binding ligand lib8 9.
KW Transforming growth factor; beta; TGF-beta; binding ligand;
KW Identification; SELEX; anti-mitogenic; inhibition; cell;
KW Systematic Evolution of Ligands by Exponential enrichment;
KW epithelial; proliferation; diagnosis; treatment; fibroids;
KW kidney; lung; liver; dermal scarring; restenosis; ss.
OS Synthetic.
PN WO9638579-A1.
PD 05-DEC-1996.
PF 30-MAY-1996; U08014.
PR 02-JUN-1995; US-458424.
PR 02-JUN-1995; US-458423.
PR 05-JUN-1995; US-465594.
PR 05-JUN-1995; US-465591.
PR 07-JUN-1995; US-479783.
PR 07-JUN-1995; US-479725.
PR 20-MAR-1996; US-618693.
PA (NEXS-) NEXSTAR PHARM INC.
PI Gold L, Janjic N, Pagratis N, Ringquist S, Toothman PJ;
DR WPI; 97-034387/03.
PT Identification of nucleic acid ligands to TGF-beta, PDGF and hKGF
PT using SELEX, used in the diagnosis and treatment of proliferative
PT disorders
PS Claim 15; Page 109; 209pp; English.
CC The present sequence, a transforming growth factor beta (TGF-beta)
CC binding ligand, was identified by Systematic Evolution of Ligands

```

CC by Exponential enrichment (SELEX). Briefly a candidate mixture of
CC nucleic acids was contacted with TGF-beta, and nucleic acids
CC having an increased affinity to TGF-beta partitioned from the
CC remainder of the mixture. The partitioned nucleic acids were then
CC amplified to yield a mixture of nucleic acids enriched for
CC sequences with higher affinity and specificity for binding to
CC TGF-beta. The ligand is anti-mitogenic and may be used to inhibit
CC epithelial cell proliferation, or in the diagnosis and treatment of
CC TGF-beta mediated pathological conditions, e.g. fibrotic conditions
CC such as fibroids of the kidney, lung and liver and more acute
CC conditions such as dermal scarring and restenosis.
SQ Sequence 123 BP; 24 A; 28 C; 45 G; 1 T; 25 U;

Query Match 69.0%; Score 13.8; DB 1; Length 123;
Best Local Similarity 64.7%; Pred. No. 77;
Matches 11; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 2 TACCTTCGGAGTATACA 18
 :|:|||||:|
Db 34 UAGCUCCGAGAGACA 50

Search completed: September 18, 1999, 05:27:15
Job time: 1680 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 06:47:28 ; Search time 1405 Seconds
(without alignments)
28.079 Million cell updates/sec

Title: US-09-037-472-2

Perfect score: 20
Sequence: 1 GTACCTCCGAGTATACATT 20

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:

- 1: em_est1.*
- 2: em_est2.*
- 3: em_est3.*
- 4: em_est4.*
- 5: em_est5.*
- 6: em_est6.*
- 7: em_est7.*
- 8: em_est8.*
- 9: em_est9.*
- 10: em_est10.*
- 11: em_est11.*
- 12: em_est12.*
- 13: em_est13.*
- 14: em_est14.*
- 15: em_est15.*
- 16: em_est16.*
- 17: em_est17.*
- 18: em_est18.*
- 19: em_est19.*
- 20: gb_est1.*
- 21: gb_est2.*
- 22: gb_est3.*
- 23: gb_est4.*
- 24: gb_est5.*
- 25: gb_est6.*
- 26: gb_est7.*
- 27: gb_est8.*
- 28: gb_est9.*
- 29: gb_est10.*
- 30: gb_est11.*
- 31: gb_est12.*
- 32: gb_est13.*
- 33: gb_est14.*
- 34: gb_est15.*
- 35: gb_est16.*
- 36: gb_est17.*
- 37: gb_est18.*
- 38: gb_est19.*
- 39: gb_est20.*
- 40: gb_est21.*
- 41: gb_est22.*
- 42: gb_est23.*
- 43: gb_est24.*
- 44: gb_est25.*
- 45: gb_est26.*
- 46: gb_est27.*
- 47: gb_est28.*
- 48: gb_est29.*
- 49: gb_est30.*
- 50: gb_est31.*
- 51: gb_est32.*
- 52: em_est20.*
- 53: em_est21.*

- 54: em_est22.*
- 55: em_est23.*
- 56: em_est24.*
- 57: em_est25.*
- 58: em_est26.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	16.8	84.0	447	29	AA183140	AA183140 mt78h05.r
C 2	15.8	79.0	187	48	AI616584	AI616584 zehno293.r
C 3	15.2	76.0	505	29	AA144931	AA144931 mr70e01.r
C 4	15.2	76.0	445	34	AA492993	AA492993 v170a10.r
C 5	15.2	76.0	392	34	AA493066	AA493066 v167h04.r
C 6	15.2	76.0	428	38	AA766803	AA766803 oa37g12.s
C 7	15.2	76.0	292	41	AI071047	AI071047 UI-R-C1-1
C 8	15.2	76.0	452	42	AI075980	AI075980 Ov47a06.x
C 9	15.2	76.0	367	46	AI414455	AI414455 mt18a02.x
C 10	15.2	76.0	149	50	AV026016	AV026016 AV026016
C 11	15.2	76.0	257	50	AV026713	AV026713 AV026713
C 12	15.2	76.0	611	51	AI728374	AI728374 BNIGH1106
C 13	15.2	76.0	625	51	AI731481	AI731481 BNIGH1987
C 14	15	75.0	614	49	AI618964	AI618964 AEMTBL29
C 15	15	75.0	635	49	AI638935	AI638935 AEMTAL76
C 16	14.8	74.0	370	21	T68771	T68771 yc29d01.r1
C 17	14.8	74.0	450	23	D52837	D52837 HUM089E02B
C 18	14.8	74.0	456	23	R93469	R93469 yq16e03.r1
C 19	14.8	74.0	337	28	C19181	C19181 C19181 Rice
C 20	14.8	74.0	339	40	AA982748	AA982748 wh15b09.r
C 21	14.8	74.0	522	43	AI239229	AI239229 GM12122.5
C 22	14.8	74.0	373	47	AI535840	AI535840 jun2.A08r
C 23	14.8	74.0	369	48	AI556178	AI556178 UI-R-C2p-
C 24	14.8	74.0	203	48	AI572265	AI572265 te39b05.x
C 25	14.4	72.0	417	20	T12971	T12971 1037 Lambda
C 26	14.4	72.0	406	23	H56099	H56099 yq95c03.r1
C 27	14.4	72.0	497	29	AA145382	AA145382 mr78g11.r
C 28	14.4	72.0	438	37	AA715483	AA715483 nv53f10.r
C 29	14.4	72.0	383	40	AA909452	AA909452 o114d11.s
C 30	14.4	72.0	593	45	AI390320	AI390320 mx03d12.y
C 31	14.4	72.0	406	46	AI449905	AI449905 mr78g11.x
C 32	14.4	72.0	528	49	AI660458	AI660458 we67d09.x
C 33	14.2	71.0	411	20	D33798	D33798 CELK036C7R
C 34	14.2	71.0	416	21	R04514	R04514 pk24f06.s1
C 35	14.2	71.0	519	21	T66346	T66346 yc78b12.r1
C 36	14.2	71.0	390	21	T77327	T77327 yc93g09.r1
C 37	14.2	71.0	354	22	H11528	H11528 ym10a02.r1
C 38	14.2	71.0	381	22	R49582	R49582 yq68g11.s1
C 39	14.2	71.0	545	22	R50886	R50886 yg70d01.r1
C 40	14.2	71.0	485	22	R51427	R51427 yg72e07.r1
C 41	14.2	71.0	437	22	R51659	R51659 yg73b04.r1
C 42	14.2	71.0	301	23	D62734	D62734 HUM320C09B
C 43	14.2	71.0	256	23	H19699	H19699 yn59h12.r1
C 44	14.2	71.0	331	23	H56379	H56379 yt87a06.r1
C 45	14.2	71.0	341	50	AV047392	AV047392 AV047392

ALIGNMENTS

RESULT 1
AA183140/c
LOCUS
DEFINITION
IMAGE:636057 5', mRNA sequence.
ACCESSION
NID
VERSION

AA183140 447 bp mRNA EST
mt78h05.r1 Soares mouse lymph node NbMLN Mus musculus CDNA clone
IMAGE:636057 5', mRNA sequence.
AA183140
g1766796
AA183140.1 GI:17666796

07-JAN-1997

KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 447)
JOURNAL Marra.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T., Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M., Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B., Theising.B., Wylie.T., Lennon.G., Soares.B., Wilson.R. and Waterston.R.
SOURCE The WashU-HMI Mouse EST Project
ORGANISM Unpublished (1996)
COMMENT On Nov 29, 1993 this sequence version replaced gi:430246.
CONTACT: Marra M/Mouse EST Project
WASHU-HMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL : contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:386049
Seq primer: -28M13 rev2 from Amerham
High quality sequence stop: 427.
Location/Qualifiers
1. 447
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:G36057"
/sex="male"
/tissue.type="lymph node"
/dev.stage="4 weeks"
/lab_host="DH10B"
/note="Organ: lymph node; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not 1; Site_2: Eco RI;
last strand cDNA was primed with a Not I - oligo(dT) primer
[5'
TGTTACCAATCGAATGGGAGCGCGGATACATTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M.Fatima Bonaldo."
BASE COUNT 154 a 77 c 101 g 115 t
ORIGIN
Query Match 84.0%; Score 16.8; DB 29; Length 447;
Best Local Similarity 90.0%; Pred. No. 21;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GACCTCCGAGTATACATT 20
Db 422 GTACTTCTAGTATACATT 403
RESULT 2
LOCUS AI616584
DEFINITION zehno293.seq.F zebrafish Embryonic Heart cDNA Library Danio rerio
ACCESSION AI616584
NID 94625751
VERSION AI616584.1 GI:4625751
KEYWORDS EST
SOURCE zebrafish.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Actinopterygii;

REFERENCE
AUTHORS Neopterygii; Teleostei; Euteleostei; Ostariophysi; Cypriniformes; Cyprinidae; Cyprinidae; Rasbora; Danio.
TITLE 1 (bases 1 to 187)
JOURNAL Ton.C., Mabry,J.D., Dempsey,A.A., Hwang,D.M., Fishman,M.C. and Liew,C.C.
COMMENT Identification and Characterization of Expressed Sequence Tags from an Embryonic Zebrafish Heart cDNA Library
Unpublished (1999)
On Jun 5, 1998 this sequence version replaced gi:3188599.
CONTACT: Liew CC
Department of Laboratory Medicine and Pathobiology
University of Toronto
Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 4169788758
Fax: 4169785650
Email: liewcc@utcc.utoronto.ca
PCR Primers
FORWARD: 5' GCCAGCTCGAATTAACCCCTCACTAAAGGG 3'
BACKWARD: 5' CCAGTGAATGTATACGACTCATATAGGCG 3'
Seq primer: 5' GAAATTAACCTCACTAAAGGG 3'.
Location/Qualifiers
1. 187
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="zebrafish Embryonic Heart cDNA Library"
/dev_stage="embryonic day 3 post-fertilization"
/lab_host="E.coli XL1-Blue mrf"
/note="Organ: heart; Vector: Lambda ZAP Express; Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from embryonic
zebrafish hearts (3 day post-fertilization). cDNA was
synthesized using a XhoI-Oligo dt adaptor-primer. EcoRI
adaptors were ligated, followed by digestion with XhoI,
for directional cloning into pre-digested lambda ZAP
Express vector."
BASE COUNT 53 a 40 c 49 g 45 t
ORIGIN
Query Match 79.0%; Score 15.8; DB 48; Length 187;
Best Local Similarity 89.5%; Pred. No. 71;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 GTACCTCCGAGTATACAT 19
Db 19 GCACCTCCGAGTATACAT 37
RESULT 3
LOCUS AA144931/c
DEFINITION mr70e01.r1 Stratagene mouse testis (#937308) Mus musculus CDNA
clone IMAGE:602808 5' similar to TR:G285949 G285949 ORF, COMPLETE
CDs. ;, mRNA sequence.
ACCESSION AA144931
NID 91714305
VERSION AA144931.1 GI:1714305
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus.
REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 505)
JOURNAL Marra.M., Hillier.L., Allen.M., Bowles.M., Dietrich.N., Dubuque.T., Geisel.S., Kucaba.T., Lacy.M., Le.M., Martin.J., Morris.M., Schellenberg.K., Steptoe.M., Tan.F., Underwood.K., Moore.B., Theising.B., Wylie.T., Lennon.G., Soares.B., Wilson.R. and Waterston.R.
SOURCE The WashU-HMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1288690.
CONTACT: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project
Washington University School
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:368240

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 336.

FEATURES

Location/Qualifiers

1..505
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:602808"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site:1:
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 Kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTT 3"

BASE COUNT 159 a 118 c 104 g 123 t 1 others

ORIGIN

Query Match 76.0%; Score 15.2; DB 29; Length 505;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTACCTTCGAGTATACATT 20

Db 128 GTCCCTGCCGAGTGATACATT 109

RESULT

AA492993/c 445 bp mRNA EST 25-JUN-1997
LOCUS vi70a10.r1 Stratagene mouse testis (#937308) Mus musculus CDNA
DEFINITION clone IMAGE:917562 5', similar to TR:G285949 G285949 ORF, COMPLETE
CDS: ., mRNA sequence.

AA492993

g2222555

AA492993.1 GI:2222555

EST.

house mouse.

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 445)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

On Sep 12, 1996 this sequence version replaced gi:1407297.

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:529778

Possible reversed clone: similarity on wrong strand

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 175.

FEATURES

Location/Qualifiers

1..445
/organism="Mus musculus"
/strain="Inbred CD-1"
/db_xref="taxon:10090"
/clone="IMAGE:917562"
/clone_lib="Stratagene mouse testis (#937308)"
/sex="males"
/tissue_type="testis"
/dev_stage="10-12 week old"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: testis; Vector: pBluescript SK-; Site:1:
EcoRI; Site:2: XhoI; Cloned unidirectionally. Primer:
Oligo dT. Average insert size: 1.0 Kb; Uni-ZAP XR Vector;
-5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor
sequence: 5' CTCGAGTTTTTTTTTTT 3"

BASE COUNT 137 a 104 c 95 g 107 t 2 others

ORIGIN

Query Match 76.0%; Score 15.2; DB 34; Length 445;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GTACCTTCGAGTATACATT 20

Db 103 GTCCCTGCCGAGTGATACATT 84

RESULT

AA493066/c 392 bp mRNA EST 25-JUN-1997
LOCUS vi67h04.r1 Stratagene mouse testis (#937308) Mus musculus CDNA
DEFINITION clone IMAGE:917335 5', similar to TR:G285949 G285949 ORF, COMPLETE
CDS: ., mRNA sequence.

AA493066

g2222628

AA493066.1 GI:2222628

EST.

house mouse.

Mus musculus

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 392)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HHMI Mouse EST Project

Unpublished (1996)

Contact: Marra M/Mouse EST Project

WashU-HHMI Mouse EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:529551

Seq primer: -28ml3 rev1 ET from Amersham

High quality sequence stop: 365.

Location/Qualifiers

1..392

/organism="Mus musculus"

/strain="Inbred CD-1"

FEATURES

Location/Qualifiers

/db_xref="taxon:10090"
 /clone="IMAGE:917335"
 /clone_lib="Stratagene mouse testis (#937308)"
 /sex="males"
 /tissue_type="testis"
 /dev_stage="10-12 week old"
 /lab_host="SOLR (kanamycin resistant)"
 /note="Organ: testis; Vector: pBluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Average insert size: 1.0 kb; Uni-ZAP XR Vector;
 -5' adaptor sequence: 5' GAATTCGGCAGAG 3' -3' adaptor
 sequence: 5' CTCGAGTTTTTTTTTTT 3'

BASE COUNT 128 a 93 c 79 g 92 t
 ORIGIN
 Query Match 76.0%; Score 15.2; DB 34; Length 392;
 Best Local Similarity 85.0%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GACCTCCGAGTATACATT 20
 ||||| ||||| |||||
 Db 87 GTCCCTCCGAGTATACATT 68

RESULT 6
 AA766803/c
 LOCUS AA766803 428 bp mRNA EST 08-FEB-1998
 DEFINITION OA37912.s1 NCI-CGAP_GCB1 Homo sapiens cDNA clone IMAGE:1307206 3'
 similar to gb:W74301.cdsl MHC CLASS II TRANSACTIVATOR CIITA
 (HUMAN); mRNA sequence.
 ACCESSION AA766803
 NID g2818041
 VERSION AA766803.1 GI:2818041
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 428)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT On Jan 19, 1998 this sequence version replaced gi:2150542.

Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman,
 Ph.D., Gerald Marti, M.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.llnl.gov/bbrp/image/image.html

Trace considered overall poor quality
 Insert Length: 914 Std Error: 0.00
 Seq primer: -40ml3 fwd. ET from Amersham
 High quality sequence stop: 1.

FEATURES

Location/Qualifiers
 1..428
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="7: 21q"
 /clone="IMAGE:1307206"
 /clone_lib="NCI-CGAP_GCB1"
 /tissue_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified

polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from human tonsillar cells enriched for
 germinal center B cells by flow sorting (CD20+, IgD-),
 provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
 (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
 primed with a Not I - oligo(dT) primer
 [5'-GGTACCAATCGAAGTGGAGCGCCGCCATTTTTTTTTTTTTT-
 3']. Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT73 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 88 a 99 c 123 g 118 t

BASE COUNT

ORIGIN
 Query Match 76.0%; Score 15.2; DB 38; Length 428;
 Best Local Similarity 85.0%; Pred. No. 1.5e+02;
 Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Oy 1 GTACCTCCGAGTATACATT 20
 ||||| ||||| |||||
 Db 243 GTACCTCCGAGTATACATT 224

RESULT 7

AI071047/c
 LOCUS AI071047 292 bp mRNA EST 11-FEB-1999
 DEFINITION UI-R-C1-10-a-10-0-UI.s1 UI-R-C1 Rattus norvegicus cDNA clone
 UI-R-C1-10-a-10-0-UI 3', mRNA sequence.
 ACCESSION AI071047
 NID 93396152
 VERSION AI071047.1 GI:3396152
 KEYWORDS EST.
 SOURCE Norway rat.
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 REFERENCE 1 (bases 1 to 292)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 9704477
 COMMENT On Jan 9, 1998 this sequence version replaced gi:936872.

Contact: Soares, MB
 Program for Rat Gene Discovery and Mapping
 University of Iowa
 451 Eckstein Medical Research Building Iowa City, IA 52242, USA
 Tel: 319 335 8250
 Fax: 319 335 9565
 Email: msoares@blue.weeg.uiowa.edu
 The sequence tag present in the cDNA between the NotI site and the
 oligo-dT track served to identify it as a clone from the normalized
 adult lung library. cDNA Library Preparation: M. Fatima Bonaldo,
 Ph.D. Clone distribution: clones will be available through Research
 Genetics The following repetitive elements were found in this cDNA
 sequence: 21-167, >LIMA4A#LINE/L1 47-246, >Lx8#LINE/L1
 Seq primer: M13 Forward.

FEATURES

Location/Qualifiers
 1..292
 /organism="Rattus norvegicus"
 /strain="Sprague-Dawley"
 /db_xref="taxon:10116"
 /clone="UI-R-C1-10-a-10-0-UI"
 /clone_lib="UI-R-C1"
 /dev_stage="Adult"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C1
 library is a subtracted library derived from the UI-R-C0
 library, which is a subtracted library derived from the

UI-R-A1 and UI-R-E1 libraries. The UI-R-A1 library consisted of a mixture of individually tagged normalized libraries constructed from rat placenta, adult lung, brain, liver, kidney, heart, spleen, ovary, and muscle. The UI-R-E1 library consisted of a mixture of individually tagged normalized libraries constructed from 8, 12 and 18-day embryo. The tag is a string of 3-5 nucleotides present between the Not I site and the oligo-dT track which allows identification of the library of origin of a clone within the mixture. The subtracted library (UI-R-C1) was constructed as follows: PCR amplified cDNA inserts from UI-R-C0 clones from which 3' ESTs had been derived was used as a driver in a hybridization with the UI-R-C0 library in the form of single-stranded circles. The remaining single-stranded circles (subtracted library) was purified by hydroxyapatite column chromatography, converted to double-stranded circles and electroporated into DH10B bacteria (Life Technologies) to generate the UI-R-C1 library. This procedure has been previously described (Bonaldo, Lennon and Soares, Genome Research 6: 791-806, 1996).

BASE COUNT 104 a 42 c 55 g 91 t
ORIGIN
Query Match 76.0%; Score 15.2; DB 41; Length 292;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GTACCTTCGAGTATACATT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 236 GTCCCTTCGTAGTATACATT 217

RESULT 8
AI075980/c
LOCUS AI075980 452 bp mRNA EST 27-AUG-1998
DEFINITION ov47a06.x1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:1640434
3', mRNA sequence.
ACCESSION AI075980
NID g3405158
VERSION AI075980.1 GI:3405158
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 452)
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tumor Gene Index
Unpublished (1997)
JOURNAL National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
COMMENT On Jan 17, 1998 this sequence version replaced gi:2045249.

Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 592 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham.
Location/Qualifiers
1. .452
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1640434"

/clone_lib="Soares_testis_NHT"
/sex="male"
/lab_host="DH10B"
/note="vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5].
TGTTACATCTGAAGTGGAGCGGCCCAATTTTTTTTTTTTTT 3'.
Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo.
BASE COUNT 136 a 90 c 94 g 132 t
ORIGIN

Query Match 76.0%; Score 15.2; DB 42; Length 452;
Best Local Similarity 85.0%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Qy 1 GTACCTTCGAGTATACATT 20
| | | | | | | | | | | | | | | | | | | | | |
Db 381 GTACCTTCGAGTATACATT 362

RESULT 9
AI141455
LOCUS AI141455 367 bp mRNA EST 09-FEB-1999
DEFINITION mb18a02.x1 Soares mouse p3NMF19.5 Mus musculus cDNA clone IMAGE:329738 3', mRNA sequence.
ACCESSION AI141455
NID g4257959
VERSION AI141455.1 GI:4257959
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 367)
AUTHORS Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
TITLE The WashU-NCI Mouse EST Project 1999
JOURNAL Unpublished (1999)
COMMENT On Apr 21, 1998 this sequence version replaced gi:3072429.

Contact: Marra M/WashU-NCI Mouse EST Project 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
This clone was previously sequenced on the 5' end only, this new data is from the 3' end
High quality sequence stop: 361.
Location/Qualifiers
1. .367
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="4p16.3; 3; X"
/clone="IMAGE:329738"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5].

FEATURES
source
Location/Qualifiers
1. .367
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="4p16.3; 3; X"
/clone="IMAGE:329738"
/dev_stage="19.5 dpc total fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Vector: pT73D (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5].

RESULT	12
AJ728374/C	
LOCUS	611 bp mRNA EST 11-JUN-1999
DEFINITION	BNLGH010609 six-day Cotton fiber Gossypium hirsutum CDNA 5' similar to (AJ000478) cytochrome P450 [Helianthus tuberosus], mRNA sequence.
ACCESSION	AJ728374
NID	G5047226
VERSION	AJ728374.1 GI:5047226
KEYWORDS	EST.
SOURCE	upland cotton. Gossypium hirsutum
ORGANISM	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium. 1. (bases 1 to 611)
REFERENCE	Blewitt.M., Matz,E.C., Davy,D.F. and Burr.B.
AUTHORS	ESRs from developing cotton fiber
TITLE	Unpublished (1999)
JOURNAL	On Jun 5, 1998 this sequence version replaced gi:3188157.
COMMENT	

```

Contact: Ben Burr
Biological Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burr@bnl.bnl.gov
Seq primer: T3 Primer.
          Location/Qualifiers
            1. .611
              /organism="Gossypium hirsutum"
              /cultivar="Acala Maxxa"
              /db_xref="taxon:3635"
              /clone_lib="Six-day Cotton fiber"
              /tissue_type="immature fiber"
              /dev_stage="Six days post anthesis"
              /lab_host="XLI-Blue"
              /note="Vector: pBluescript II KS+"
BASE COUNT      202 a      105 c      136 g      168 t
ORIGIN

Query Match          76.08;   Score 15.2;   DB 51;   Length 611;
Best Local Similarity 85.08;   Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

QY	1	GTACCTTCGGAGTAATACATT 20
Dd	569	GTACCTTCCCCTAAACATT 550
RESULT	13	
LOCUS	A1731481/c	
DEFINITION	A1731481 625 bp mRNA EST BNLHJ1979 Six-day Cotton fiber Gossypium hirsutum cDNA 5' similar to (AJ000478) cytochrome P450 [Helianthus tuberosus], mRNA sequence.	
ACCESSION	A1731481	
NID	G5050333	
VERSION	A1731481.1 GI:5050333	
KEYWORDS	EST.	
SOURCE	upland cotton.	
ORGANISM	Gossypium hirsutum	
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Malvales; Malvaceae; Gossypium. 1 (bases 1 to 625)	
AUTHORS	Blevitt,M., Matz,E.C., Davy,D.F. and Burr,B.	
TITLE	ESTs from developing cotton fiber	
JOURNAL	Unpublished (1999)	
COMMENT	On Mar 10, 1998 this sequence version replaced gi:2948292.	

```

Contact: Ben Burr
Biology Department
Brookhaven National Laboratory
Upton, NY 11973, USA
Tel: 516-344-3396
Fax: 516-344-3407
Email: burr@bnl.bnl.gov
Seq primer: T3 Primer.
Location/Qualifiers
1. .625
/organism="Gossypium hirsutum"
/cultivar="Acala Maxxa"
/db_xref="taxon:3635"
/clone_lib="Six-day Cotton fiber"
/tissue_type="immature fiber"
/dev_stage="Six days post anthesis"
/lab_host="XL1-Blue"
/note="vector: pBluescript II KS+"
BASE COUNT      208 a   109 c   138 g   170 t
ORIGIN

Query Match      76.08;   Score 15.2;   DB 51;   Length 625;
Best Local Similarity 85.08;   Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 GTACCTCCGAGTATACATT 20
      ||||| ||| || |||||
Db 569 GTACCTCCCACTAAACATT 550

RESULT 14
A1618964
LOCUS
DEFINITION A1618964 614 bp mRNA EST 21-APR-1999
AEMTB129 Aedes aegypti MT pSPORT Library Aedes aegypti cDNA clone
B129 5', mRNA sequence.
A1618964
NTL 4628090
VERSION A1618964.1 GI:4628090
KEYW KDS EST.
SOURCE yellow fever mosquito.
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Culicidae; Aedes.
1 (bases 1 to 614)
Gill,S.S., Ross,L.S. and Wadiak,H.
Expressed sequence tags of cDNA clones from an enriched Malpighian
tubule and gut library from Aedes aegypti
Unpublished (1999)
JOURNAL On Mar 10, 1998 this sequence version replaced gi:2948045.
COMMENT

```

```

Contact: Gill SS
Environmental Toxicology
University of California, Riverside
5419 Boyce Hall, Riverside, CA 92521, USA
Tel: 909 787 3547
Fax: 909 787 3087
Email: sarjeet.gill@ucr.edu
Seq primer: CCAAGCTCAATACGACTCACTAT
High quality sequence stop: 600.
Location/Qualifiers
1..614
/organism="Aedes aegypti"
/db_xref="taxon:7159"
/clone="BL29"
/clone_lib="Aedes aegypti MT pSPORT Library"
/sex="female"
/tissue_type="malpighian tubules and gut"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; The

```

cdna was cloned into the Sali/NotI sites of pSPORT1."

BASE COUNT 158 a 110 c 162 g 184 t
ORIGIN

Query Match 75.0%; Score 15; DB 49; Length 614;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TTCCGAGTATACATT 20
|||||
Db 357 TTCCGAGTATACATT 371

RESULT 15
AI638935
LOCUS A1638935 635 bp mRNA EST 27-APR-1999
DEFINITION AEMTAL76 Aedes aegypti MT pSPORT Library Aedes aegypti cDNA clone
AL76.5', mRNA sequence.
ACCESSION A1638935
NID 94691169
VERSION A1638935.1 GI:4691169
KEYWORDS EST.
SOURCE yellow fever mosquito.
ORGANISM Aedes aegypti
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Nematocera;
Culicoidae; Culicidae; Aedes.
REFERENCE Gill, S.S., Ross, L.S. and Wadiak, H.
AUTHORS 1 (bases 1 to 635)
TITLE Expressed sequence tags of cDNA clones from an enriched Malpighian
tubule and gut library from Aedes aegypti
JOURNAL Unpublished (1999)
COMMENT On Mar 16, 1998 this sequence version replaced gi:2961689.

Contact: Gill SS
Environmental Toxicology
University of California, Riverside
5419 Boyce Hall, Riverside, CA 92521, USA
Tel: 909 787 3547
Fax: 909 787 3087
Email: sarjeet.gill@ucr.edu
Seq primer: CCAAGCTCTAATACGACTCACTAT
High quality sequence stop: 635.

FEATURES
source
1..635
/organism="Aedes aegypti"
/db_xref="taxon:7159"
/clone="AL76"
/clone_lib="Aedes aegypti MT pSPORT Library"
/sex="female"
/tissue_type="malpighian tubules and gut"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: Sali; Site_2: NotI; The
cdna was cloned into the Sali/NotI sites of pSPORT1."
BASE COUNT 160 a 116 c 169 g 190 t
ORIGIN

Query Match 75.0%; Score 15; DB 49; Length 635;
Best Local Similarity 100.0%; Pred. No. 2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TTCCGAGTATACATT 20
|||||
Db 373 TTCCGAGTATACATT 387

Search completed: September 18, 1999, 06:47:32
Job time: 3649 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model
Run on: September 18, 1999, 15:25:20 : Search time 436.05 seconds
(without alignments)
196.923 Million cell updates/sec

Title: US-09-037-472-1
Perfect score: 27
Sequence: 1 AAGCTTCTTACCACTGAAGTAGGC 27

Scoring table: IDENTITY_NUC
Searched: 679419 seqs, 1590154680 residues

Database :

GenEmbl.*

1: gb_ba1.*

2: gb_ba2.*

3: gb_ov.*

4: gb_ov.*

5: gb_ph.*

6: gb_ph.*

7: gb_pl1.*

8: gb_pl2.*

9: gb_pr1.*

10: gb_pr2.*

11: gb_pr3.*

12: gb_ro.*

13: gb_st.*

14: gb_sts.*

15: gb_sy.*

16: gb_un.*

17: gb_vl.*

18: em_fun.*

19: em_hg.*

20: em_hum1.*

21: em_hum2.*

22: em_in.*

23: em_or.*

24: em_or.*

25: em_ov.*

26: em_pat.*

27: em_ph.*

28: em_pl.*

29: em_ro.*

30: em_sts.*

31: em_sy.*

32: em_un.*

33: em_vl.*

34: gb_hg1.*

35: gb_hg2.*

36: gb_in1.*

37: gb_in2.*

38: gb_ba1.*

39: gb_ba2.*

40: em_hum3.*

41: em_hum4.*

42: gb_pr4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	DB	ID	Description
1	22	81.5	22	5	I73225	Sequence 1

2	22	81.5	11970	10	HSILIAG	X03833 Human gene
3	18.6	68.9	145614	34	AC006804	AC006804 Caenorhab
4	18.6	68.9	145614	34	AC006872	AC006872 Caenorhab
5	18.6	68.9	244239	34	AC006880	AC006880 Caenorhab
6	18.6	68.9	151528	35	AC006447	AC006447 Mus muscu
c 7	18.4	68.1	170952	9	HS01N12	AL022170 Homo sapi
c 8	18.2	67.4	103574	10	HSAC002115	AC002115 Human DNA
c 9	18	66.7	1910	4	CHLNKPA2	M13212 Chicken car
c 10	18	66.7	460	4	CHLNKPA2	M35036 Chicken car
c 11	18	66.7	14313	7	NEURO1DHC	L31504 Neurospora
c 12	18	66.7	139166	11	AC003029	AC003029 Homo sapi
c 13	18	66.7	151795	34	HS85661	AL033361 Homo sapi
c 14	18	66.7	51050	36	AC004340	AC004340 Drosophil
c 15	18	66.7	85139	37	AC001657	AC001657 Drosophil
c 16	18	66.7	196672	37	AC004767	AC004767 Drosophil
17	17.6	65.2	16094	1	MTV038	AL021933 Mycobacte
18	17.6	65.2	783	3	AF030474	AF030474 Gazella s
19	17.6	65.2	783	3	AF030475	AF030475 Gazella c
20	17.6	65.2	783	3	AF030476	AF030476 Gazella l
21	17.6	65.2	783	3	AF030477	AF030477 Gazella s
22	17.6	65.2	783	3	AF030478	AF030478 Gazella s
23	17.6	65.2	783	3	AF030479	AF030479 Gazella b
24	17.6	65.2	783	3	AF030480	AF030480 Gazella b
25	17.6	65.2	783	3	AF030481	AF030481 Gazella s
26	17.6	65.2	783	3	AF030483	AF030483 Gazella d
27	17.6	65.2	783	3	AF030484	AF030484 Gazella d
28	17.6	65.2	783	3	AF030485	AF030485 Gazella g
29	17.6	65.2	783	3	AF030487	AF030487 Gazella g
30	17.6	65.2	783	3	AF030488	AF030488 Gazella g
31	17.6	65.2	783	3	AF030489	AF030489 Gazella g
32	17.6	65.2	16034	34	AC006812	AC006812 Caenorhab
33	17.6	65.2	16034	34	AC006812	AC006812 Caenorhab
c 34	17.6	65.2	90143	34	HSAC001232	AC001232 Homo sapi
35	17.6	65.2	171336	35	AC007255	AC007255 Homo sapi
c 36	17.6	65.2	473	36	AB002162	AB002162 Entamoeba
c 37	17.4	64.4	6081	7	ALCRDNA	X14386 Astasia lon
38	17.4	64.4	37000	7	SPAC19G12	297209 S.pombe chr
39	17.4	64.4	3057	8	AF038122	AF038122 Podospora
c 40	17.4	64.4	1542	9	HUMAOP1	D49396 Human mRNA
c 41	17.4	64.4	115973	11	AC004823	AC004823 Homo sapi
c 42	17.4	64.4	92611	11	AC006398	AC006398 Homo sapi
c 43	17.4	64.4	184514	34	AC006466	AC006466 Homo sapi
c 44	17.4	64.4	39897	34	LMFL1156	AL034390 Leishmani
c 45	17.4	64.4	167713	35	AC007778	AC007778 Homo sapi

ALIGNMENTS

RESULT 1	I73225	Sequence 1 from patent US 5686246.	PAT	23-DEC-1997
LOCUS	I73225	22 bp DNA		
DEFINITION	Sequence 1 from patent US 5686246.			
ACCESSION	I73225			
NID	93009364			
VERSION	I73225.1	GI:3009364		
KEYWORDS				
SOURCE	Unknown.			
ORGANISM	Unknown.			
REFERENCE	Unclassified.			
AUTHORS	1 (bases 1 to 22)			
TITLE	Kornman, K.S. and Duff, G.W.			
JOURNAL	Detecting genetic predisposition to periodontal disease			
FEATURES	Patent: US 5686246-A 1 11-NOV-1997;			
	Location/Qualifiers			
	1..22			
BASE COUNT	5 a	7 c	4 g	6 t
ORIGIN	/Organism="unknown"			

Query Match 81.5% Score 22; DB 5; Length 22;
Best Local Similarity 100.0%; Pred. No. 2.2;

Query Match	81.5%	Score 22;	DB 10;	Length 11970;
Best Local Similarity	100.0%;	Pred. No. 2.1;		
Matches 22;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy 6	TGTTCTACCACTGAACTAGGC 27			
Db 471	TGTTCTACCACTGAACTAGGC 492			
RESULT 3				
AC006804				
LOCUS	AC006804 145614 bp DNA HTG			25-FEB-1999
DEFINITION	Caenorhabditis elegans clone Y53G8B, WORKING DRAFT SEQUENCE, 5 unordered pieces.			
ACCESSION	AC006804			
NID	94309782			
VERSION	AC006804.2 GI:4309782			
KEYWORDS	HTG; HTGS_PHASE1.			
SOURCE	Caenorhabditis elegans.			
ORGANISM	Caenorhabditis elegans			
REFERENCE	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae; Rhabditina; Rhabditoidea; Rhabditoidea; Peloderinae; Caenorhabditis.			
AUTHORS	1 (bases 1 to 145614)			
TITLE	Waterston, R.H.			
JOURNAL	The sequence of Caenorhabditis elegans clone			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 145614)			
TITLE	Waterston, R.H.			
JOURNAL	Direct Submission			
COMMENT	Submitted (23-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA			
	On Mar 1, 1999 this sequence version replaced gi:4263134.			
	* NOTE: This is a 'working draft' sequence. It currently			
	* consists of 5 contigs. The true order of the pieces			
	* is not known and their order in this sequence record is			
	* arbitrary. Caenorhabditis elegans			

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

1 19062: contig of 19062 bp in length
19063 19071: gap of unknown length
19072 27630: contig of 3559 bp in length
27631 27639: gap of unknown length
27640 54395: contig of 25756 bp in length
54396 54404: gap of unknown length
54405 56414: contig of 2010 bp in length
56415 56423: gap of unknown length
56424 145614: contig of 89191 bp in length.

Location/Qualifiers
1. .145614
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y53G8B"

BASE COUNT 46966 a 25460 c 25159 g 47086 t 943 others
ORIGIN

Query Match 68.9%; Score 18.6; DB 34; Length 145614;
Best Local Similarity 84.0%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTTACCACCTGAAGTAG 25
| |||| ||||||||| ||| |
Db 53675 ATGCTTCTTCTACCACTGAAGATTG 53699

FEATURES
source
1. .145614
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y53G8B"

BASE COUNT 46966 a 25460 c 25159 g 47086 t 943 others
ORIGIN

Query Match 68.9%; Score 18.6; DB 34; Length 145614;
Best Local Similarity 84.0%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTTACCACCTGAAGTAG 25
| |||| ||||||||| ||| |
Db 53675 ATGCTTCTTCTACCACTGAAGATTG 53699

RESULT 4
AC006872 145614 bp DNA HTG 26-FEB-1999
LOCUS Caenorhabditis elegans clone Y53G8Y, WORKING DRAFT SEQUENCE, 5
DEFINITION unordered pieces.
ACCESSION AC006872
NID G4309781
VERSION AC006872.2 GI:4309781
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 145614)
AUTHORS Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
TITLE Waterston,R.H.
JOURNAL The sequence of Caenorhabditis elegans clone
REFERENCE 2 (bases 1 to 145614)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (24-FEB-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 19062: contig of 19062 bp in length
19063 19071: gap of unknown length
19072 27630: contig of 8559 bp in length
27631 27639: gap of unknown length
27640 54395: contig of 26756 bp in length
54396 54404: gap of unknown length
54405 56414: contig of 2010 bp in length
56415 56423: gap of unknown length
56424 145614: contig of 89191 bp in length.

Location/Qualifiers

source
1. .145614
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y53G8Y"

BASE COUNT 46966 a 25460 c 25159 g 47086 t 943 others
ORIGIN

Query Match 68.9%; Score 18.6; DB 34; Length 145614;
Best Local Similarity 84.0%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTTACCACCTGAAGTAG 25
| |||| ||||||||| ||| |
Db 53675 ATGCTTCTTCTACCACTGAAGATTG 53699

RESULT 5
AC006880 244239 bp DNA HTG 24-FEB-1999
LOCUS Caenorhabditis elegans clone Y55D5, WORKING DRAFT SEQUENCE, 4
DEFINITION unordered pieces.
ACCESSION AC006880
NID G4263468
VERSION AC006880.1 GI:4263468
KEYWORDS HTG; HTGS_PHASE1.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.
REFERENCE 1 (bases 1 to 244239)
AUTHORS Waterston,R.H.
JOURNAL The sequence of Caenorhabditis elegans clone
TITLE Unpublished
REFERENCE 2 (bases 1 to 244239)
AUTHORS Waterston,R.H.
JOURNAL Direct Submission
TITLE Submitted (24-FEB-1999) Genome Sequencing Center, Washington
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
COMMENT * NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 9998: contig of 9998 bp in length
9999 10012: gap of unknown length
10013 19150: contig of 9138 bp in length
19151 19164: gap of unknown length
19165 88195: contig of 69031 bp in length
88196 88209: gap of unknown length
88210 244239: contig of 156030 bp in length.

Location/Qualifiers
1. .244239
/organism="Caenorhabditis elegans"
/db_xref="taxon:6239"
/clone="Y55D5"

BASE COUNT 78698 a 43791 c 42693 g 79015 t 42 others
ORIGIN

Query Match 68.9%; Score 18.6; DB 34; Length 244239;
Best Local Similarity 84.0%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTTACCACCTGAAGTAG 25
| |||| ||||||||| ||| |
Db 87475 ATGCTTCTTCTACCACTGAAGATTG 87499

```

RESULT 6
AC006447  AC006447  151528 bp  DNA  HTG  09-JUN-1999
LOCUS      Mus musculus, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION  AC006447
NID        95030453
VERSION    HTG: HTGS-PHASE1.
KEYWORDS   house mouse.
SOURCE     Mus musculus
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 151528)
AUTHORS    Wu,H., Yao,Z., McDermid,H. and Roe,B.A.
TITLE      Mus musculus Chromosome 6 BAC Clone 67d14
JOURNAL    Unpublished
REFERENCE  2 (bases 1 to 151528)
AUTHORS    Wu,H., Yao,Z., McDermid,H. and Roe,B.A.
TITLE      Direct Submission
JOURNAL    Submitted (28-JAN-1999) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT    On Jun 9, 1999 this sequence version replaced gi:4580404.
            * NOTE: this is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1 5480: contig of 5480 bp in length
            * 5481 5555: gap of unknown length
            * 12748: contig of 7193 bp in length
            * 5556 12748: gap of unknown length
            * 12749 12823: gap of unknown length
            * 12824 37024: contig of 24201 bp in length
            * 37025 37099: gap of unknown length
            * 37100 82029: contig of 44930 bp in length
            * 82030 82104: gap of unknown length
            * 82105 151528: contig of 69424 bp in length.
FEATURES   Location/Qualifiers
            1..151528
               /organism="Mus musculus"
               /db_xref="taxon:10090"
BASE COUNT 43172 a 33398 c 33443 g 41215 t 300 others
ORIGIN
Query Match 68.9%; Score 18.6; DB 35; Length 151528;
Best Local Similarity 84.0%; Pred. No. 79;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 AAGCTTGTTCTACCACTGAACCTAG 25
Db 98144 AAGCTTGTTCTACCACTGAACCTAG 98168
|||||
|||||

RESULT 7
HS501N12/c  HS501N12  170952 bp  DNA  PRI  24-JUL-1998
LOCUS      Homo sapiens DNA sequence from clone 501N12 on chromosome
DEFINITION 6p22.1-22.3 Contains a gene almost identical to four genes of
            unknown function, a pseudogene, three (pseudot?) genes similar to
            genes of unknown function, an unknown gene similar to a rat EST, a
            PX19 LIKE pseudogene and another unknown gene. Contains ESTs, STS
            and GSSs, complete sequence.
ACCESSION  AL022170
NID        g3281976
VERSION    AL022170.1 GI:3281976
KEYWORDS   HTG: PX19.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;

```

```

REFERENCE 1 (bases 1 to 170952)
AUTHORS   Tubby,B.
TITLE     Direct Submission
JOURNAL   Submitted (06-JUL-1998) E-mail enquires: humquery@sanger.ac.uk
COMMENT   On Jul 1, 1998 this sequence version replaced gi:2980811.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above. This sequence is the
            entire insert of clone 501N12. This sequence has been finished
            according to sequence map criteria as follows. An attempt is made
            to resolve all sequencing problems, such as compressions and
            repeats, but not necessarily within known annotated human repeat
            sequence elements (e.g. Alu). Where the sequence is ambiguous,
            there is an annotation using the 'unsure' feature key.
            This sequence was generated from part of bacterial clone contigs of
            human chromosome 6, constructed in collaboration by the Sanger
            Centre chromosome 6 mapping group and Armin Volz & Andreas Ziegler.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6/501N12 is from the library RPC13
            constructed at the Roswell Park Cancer Institute by the group of
            Pieter de Jong. For further details see
            http://bacpac.med.buffalo.edu/VECTOR: pCYPAC2.
FEATURES   Location/Qualifiers
            1..170952
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="6"
               /clone="501N12"
               /map="p22.1-22.3"
               /clone_lib="RPC13"
               444..516
               /note="MER5A repeat: matches 101..188 of consensus"
               complement(832..1100)
               /note="AluJb repeat: matches 280..1 of consensus"
               complement(1341..1477)
               /note="MIR repeat: matches 237..85 of consensus"
               2942..3479
               /note="L1 repeat: matches 1432..1968 of consensus"
               4030..4359
               /note="AluJ repeat: matches 1..296 of consensus"
               complement(4554..4853)
               /note="AluJo repeat: matches 301..1 of consensus"
               6829..34852
               /gene="dJ501N12.1"
               join(<6829..6895,15857..15985,17949..18080,28912..29087,
               32011..32098,34703..34852)
               /gene="dJ501N12.1"
               /note="match: cDNAs Y10518 Y10519 Z83950 Y10508"
               /codon_start=2
               /evidence=not_experimental
               /product="dJ501N12.1"
               /protein_id="CAA18156.1"
               /db_xref="PID:e1312796"
               /db_xref="PID:g3355565"
               /db_xref="GI:3355565"
               /db_xref="SPTREMBL:O75660"
               /translation="RYLQEQRYRLOQIVTSTQQMIDRICVKVDHLNLRNCGGD
               AIOEDLKSERLMDARNKSTLLPNLYHVGASWAGSLSSPIQETLSEMAGEVTR
               VVDEQKALLSEVDAENLCPNMKAHIRQDLIHASTEKISIPRTFVKVNLLEQSG
               IDLTKISVEKLVTAFLSDRIVDEILDALSHCHHKLADHFSRRGKTLPQGESLEIEL
               AEEKPKRSIITVEELTEERLEDLTCM"
               7271..7658
               /gene="dJ501N12.1"
               /note="match: EST AA526622"
               7659..7776
               /note="AluYb8 repeat: matches 191..308 of consensus"
               8239..8475
               /note="AluSg repeat: matches 1..239 of consensus"

```


TITLE Direct Submission
 JOURNAL Submitted (13-MAR-1997) Human Genome Center, Lawrence Livermore
 National Laboratory, 7000 East Ave., Livermore, CA 94551, USA
 COMMENT R31396 from 1- 36,162; F25451 from 25,661-55,793; R31076 from
 66,237- 103,574.
 FEATURES Location/Qualifiers
 source 1..103574
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="R31396-F25451-R31076"
 /chromosome="19"
 /map="19q13.1 between D19S208 and CAPNS"
 /map="overlaps CH19F14121 to the left and CH19R28052 to
 the right"
 /cell_type="fibroblast"
 /map="orientation is centromere to telomere"
 /note="Cosmid libraries constructed at LLNL from
 flow-sorted chromosomes from hybrids UV5HL9-5B and 5HL2-B,
 which carry chromosome 19 as their only human chromosome"
 complement(187..466)
 /rpt_family="ALU"
 /rpt_family="L1"
 719..842
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 2, quality: good, score: 57.000"
 complement(957..1296)
 /rpt_family="MER41"
 1289..1590
 /rpt_family="ALU"
 1907..2112
 /rpt_family="L1"
 complement(2199..2280)
 /rpt_family="ALU"
 complement(2310..2392)
 /rpt_family="ALU"
 2377..2663
 /rpt_family="ALU"
 3085..3341
 /rpt_family="ALU"
 3429..3705
 /rpt_family="ALU"
 <3755..4205
 /note="BLASTX similarity to (283..429); match: 0.47,
 score: 5.7e-29; database searched: nr; hypothetical L1
 protein (third intron of gene TS)- human >prf11510254A L1
 repetitive element ORF [Homo sapiens]"
 /rpt_family="L1"
 complement(4223..4455)
 /rpt_family="ALU"
 complement(4517..4584)
 /rpt_family="ALU"
 complement(4822..4931)
 /rpt_family="MIR"
 5127..5417
 /rpt_family="ALU"
 complement(5730..6009)
 /rpt_family="ALU"
 complement(7534..7834)
 /rpt_family="ALU"
 complement(8629..8915)
 /rpt_family="ALU"
 9070..9183
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 0, quality: excellent, score: 94.000"
 10193..10276
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 1, quality: good, score: 56.000"
 11501..11753
 /rpt_family="ALU"
 11801..12092
 /rpt_family="ALU"
 complement(12570..12861)
 /rpt_family="ALU"
 12958..13233

repeat_region /rpt_family="ALU"
 complement(13377..13672)
 repeat_region /rpt_family="ALU"
 13899..14179
 /rpt_family="ALU"
 14234..14310
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 0, quality: good, score: 56.000"
 14660..14961
 /rpt_family="ALU"
 14946..15122
 /rpt_family="L1"
 15173..15445
 /rpt_family="ALU"
 15584..15619
 /rpt_family="ALU"
 complement(15626..15915)
 /rpt_family="ALU"
 complement(15950..16258)
 /rpt_family="ALU"
 16665..17493
 /rpt_family="ALU"
 17507..18084
 /rpt_family="ALU"
 complement(18269..18375)
 /rpt_family="MER21"
 19017..19084
 /rpt_family="ALU"
 19245..19346
 /rpt_family="ALU"
 complement(19351..19626)
 /rpt_family="ALU"
 19697..19833
 /rpt_family="ALU"
 20425..20701
 /rpt_family="ALU"
 complement(20717..20970)
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 1, quality: marginal, score: 42.000"
 21268..21858
 /rpt_family="ALU"
 21922..22503
 /rpt_family="ALU"
 22690..22803
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 0, quality: good, score: 74.000"
 complement(22986..23280)
 /rpt_family="ALU"
 23377..23497
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 1, quality: good, score: 57.000"
 complement(23687..23971)
 /rpt_family="ALU"
 24001..24169
 /rpt_family="MER21"
 24305..24583
 /rpt_family="ALU"
 25093..25377
 /rpt_family="ALU"
 complement(25389..25668)
 /rpt_family="MER31"
 25459..25516
 /note="Predicted exon, program: grail2exons_human_1.3,
 frame: 0, quality: excellent, score: 79.000"
 complement(25784..26071)
 /rpt_family="ALU"
 complement(26118..26558)
 /rpt_family="ALU"
 26645..27013
 /rpt_family="LTR7"
 27474..27765
 /rpt_family="ALU"
 complement(28025..28395)
 repeat_region

/rpt_family="THE1"
complement(28597..28876)
/rpt_family="ALU"
complement(29166..29667)
/rpt_family="MER9"
complement(29979..30276)
/rpt_family="ALU"
30918..31090
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 95.000"
31106..31666
/rpt_family="ALU"
complement(31848..31968)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: marginal, score: 45.000"
complement(32692..32734)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 94.000"
33296..33518
/rpt_family="MER1"
33532..33829
/rpt_family="ALU"
33885..33983
/rpt_family="MER1"
complement(34062..34170)
/note="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 97.000"
34376..34477
/rpt_family="ALU"
35483..35637
/rpt_family="ALU"

Query Match 67.4%; Score 18.2; DB 10; Length 103574;
Best Local Similarity 87.0%; Pred. No. 1.2e+02;
Matches 20; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 4 CTTGTTCTTACCCTGAACTAGG 26
||||||| || |||||
Db 22876 CTTGTTCTCCCTTCTGAACTAGG 22898

RESULT 9
CHKLKPA/c
LOCUS CHKLKPA 1910 bp mRNA VRT 04-FEB-1991
DEFINITION Chicken cartilage link protein mRNA, complete cds.
ACCESSION M13212
NID 9212259
VERSION M13212.1 GI:212259
KEYWORDS
SOURCE
ORGANISM *Chicken embryo (14 day) sternal cartilage, cDNA to mRNA.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 1910)
Deak,F., Kiss,I., Sparks,K.J., Argraves,W.S., Hampikian,G. and
Goetinck,P.F.
Complete amino acid sequence of chicken cartilage link protein
deduced from cDNA clones
Proc. Natl. Acad. Sci. U.S.A. 83, 3766-3770 (1986)
86233315
FEATURES
Location/Qualifiers
1..1910
/organism="Gallus gallus"
/db_xref="taxon:9031"
136..1203
/note="cartilage link protein"
/codon_start=1
/protein_id="AA48940.1"
/db_xref="PID:g212260"
/db_xref="GI:212260"
/translation="MTSLFLVLISVCWAEPHPDSSLEHRIIHQEEGPRLLVVA

EQAKIFSQRGGNVTLPCKFYHEHTSTAGSGTHKIRVWKLTSDYLKEVDVFVAMGHH
RKSXYGRGVFLRESSNDASLIITNIMEDYGRYKCEVIEGLEDDTAVVALNLEGV
VFYSPRLGRYNLNEHQAQCLDSDSIITASEDLYEAMRSGLDWAGWLSGVSQY
PTKPREPGCGKNTVPGVNYGFWDKERSYDVFCITSNFGRTYLIIPTKLTYDEA
VQACLDGAQIAKVGQIFAAWKLGLGYDRCDAGLADGSVRYPIISRKRKCSPPNEAAVR
FVGFPDKKHLXGVYCFRAYN"
BASE COUNT 594 a 389 c 408 g 519 t
ORIGIN

Query Match 66.7%; Score 18; DB 4; Length 1910;
Best Local Similarity 80.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AGCTTGTTCTTACCCTGAACTAGGC 27
||||||| || |||||
Db 276 AGCTTGTTCTTACCACAAAGTAGGC 251

RESULT 10
CHKLKPA2/c
LOCUS CHKLKPA2 460 bp DNA VRT 15-SEP-1990
DEFINITION Chicken cartilage link protein gene, exon 3.
ACCESSION M35036
NID 9212262
VERSION M35036.1 GI:212262
KEYWORDS cartilage link protein.
SEGMENT 2 of 5
SOURCE Chicken (domesticus, strain White Leghorn) 9-day embryo DNA, clones
lambda gLP39.13 and lambda gLP33.7.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Archosauria;
Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 460; 49 to 63; 429 to 442)
Kiss,I., Deak,F., Mestric,S., Delius,H., Soos,J., Dekany,K.,
Argraves,W.S., Sparks,K.J. and Goetinck,P.
Structure of the chicken link protein gene: Exons correlate with
the protein domains
Proc. Natl. Acad. Sci. U.S.A. 84, 6399-6403 (1987)
87317659
COMMENT Draft entry and computer-readable sequence for [1] kindly submitted
by I.Kiss, 04-JUN-1990.
FEATURES
Location/Qualifiers
1..460
/organism="Gallus gallus"
/db_xref="taxon:9031"
prim_transcript <1..>460
intron <1..58
/note="cartilage link protein mRNA and introns"
exon 59..433
/number=3
intron 434..460
/note="cartilage link protein intron C"
BASE COUNT 145 a 90 c 113 g 112 t
ORIGIN

Query Match 66.7%; Score 18; DB 4; Length 460;
Best Local Similarity 80.8%; Pred. No. 1.6e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 AGCTTGTTCTTACCCTGAACTAGGC 27
||||||| || |||||
Db 99 AGCTTGTTCTTACCACAAAGTAGGC 74

RESULT 11
NEUROIDHC
LOCUS NEUROIDHC 14313 bp DNA PLN 07-APR-1995
DEFINITION Neurospora crassa cytoplasmic dynein heavy chain (ro-1) gene,
complete cds.
ACCESSION L31504

9473489
 L31504.1 GI:473489
 cytoplasmic dynein heavy chain.
 Neurospora crassa DNA.
 Neurospora crassa
 Eukaryota; Fungi; Ascomycota; Euascomycetes; Pyrenomycetes;
 Sordariales; Sordariaceae; Neurospora.
 1 (bases 1 to 14313)
 Plamann,M., Minko,P.F., Tinsley,J.H. and Bruno,K.S.
 Cytoplasmic dynein and actin-related protein Arp1 are required for
 normal nuclear distribution in filamentous fungi
 J. Cell Biol. 127 (1), 139-149 (1994)
 95014704
 Location/Qualifiers
 1. 14313
 /organism="Neurospora crassa"
 /db_xref="taxon:5141"
 join(173..482,549..12853,12926..13414)
 /gene="ro-1"
 join(173..482,549..12853,12926..13414)
 /gene="ro-1"
 /codon_start=1
 /product="cytoplasmic dynein heavy chain"
 /protein_id="AAA64508.1"
 /db_xref="PID:9473490"
 /db_xref="GI:473490"
 /translation="MMDSPSPPPQSPDANGVATTFFAADVPKVVVDHLVLLLEATL
 GAKRDEAPGSLLSKRVYSDTVQRCRFALDQTOVALYIQKOLAPTITLIDGNGAEE
 EPEPTHTVYISDLSSTSPYAVLVLLKRPOLDPIVPLTSIQMLNLPGATLSTSG
 SEQGPSPEYELQLKXGLAPYFDASTKSOQLNGARGVDQAKTGIPVTKKMT
 ELESUSHLOQNVIEPVSPLPHVQSQTELEAATKNVPSIDLLPALTADSTFLNN
 LQATVNWQISQVITKMDRPTTGTANQIEINFWLSMEAALEGIEINQLRSEGVMLLN
 ILKHKARFOATFTADPLGKAMEKVQYNOLMRODFPLDELLSATLTKVQESIGQI
 LKHNKKLRCPVPIRRALPLVEISGDLDEVLHLLPGTELVKLDIEFEFGVMKQAG
 SIFRWDESIKEFTNVAREVSTRNRNEKFTPIKINPHAEQLQSRLDTHVHFRHNEQL
 RIINVLPKATVNGVLTASGAVVEIGDVDAVDEKQAKDLVLDLDDCTRE
 GIEKVRAGENINERTAVENSIIARLDRLATKANEMFVSKFNALFVRPKIRG
 LAEYQTDIDNVKQASISLHFQKQYSEAHAAQLHDLPPVSGAIIWAKOISRG
 LDQMKYQGVQSLQSWALTEGQKQNESDLFRKKLDTRPIFEAWLHDVQRKQIESQ
 LLFTINIRISAGNILEAVNFOAVTALFKETNLLNLNYPVPHSVNNVAKKARYVP
 FAVLSMVRTAQTNRQISDSEVAVLTSRGNNDVYTLISKGIPLRWYVYQNTLEH
 EPTFPNPTPLQGSKVSETHVMFIREFAASLLQSKTLILLANIYVTKVLEVR
 KICYPASAFQRLETIQHAVDNLQEQVNLGYWYVRNMQIKDLYTRLQVIAQW
 IOAFEDVERSEKRLLEIASPOAKSIGVIKSLVHEITMRQVIVLDPPELVAR
 ASWAFQLDWIGVGNLKKIKATRYMSTLSTVEDPREFNDLPQCTELLRVQTSVE
 KIRIEGVAGWKWLPQSPVQSEHVEDVGLDSRWLQLOEIRKTRQETDTIYS
 RSEGHITIDYDVGIKVNAKYVQWODILIKFASRLGNRMREYALEKARKDLEQA
 MTANSTAEAVREITVACTROYKWLPEITETPROGESTLVQRHFQNDMLHAEQV
 GAWMDNELLEAKRSKIVDOOALRAKRTIAEDKVNDKIAEIAHWNEEKPVSGTIAP
 DVASATLFEQIRTKQELSAMAKAEALDAPTDPTSLGVLIEEVDPKSVWML
 STIKNLNRELTWNVSVPRIYASAKIDNLKMTKEMPSRMROYAAFEHIQNVLRQLM
 KQVGMDELLEKSEAVDRHWKIKYQKPKGKYSPVSMTLGDVNDLNLVATEVIXDII
 IOAGSMALKEELFKOVRETTWYGLIELVQYQKRLIRGWDOLFACSNLSLQAKM
 HSPYKFEPEEASSEEKLNRVHVLFDIWDVQRYWYLEGVHGNADNLHLLPTES
 RFQNIINSEFLVAMKVIYKOPNVLDVNLINPVQKSLRSLAEALNKKIQALGEYLEKRV
 SPFRYFVGEDLLENGIMNTMRIAKHFKKMFAGLNGLVMDGEVIGSGFTSKEGET
 VLRKLEINLVKTRPINDMLALENGKMTYLAELLAQVEDFTPIFSESNVDRDALIKF
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIGIQVGWAGCFDEFNRLEEILSAVSQOIQDIO
 LGLNGAEDEKAQIILDGRQIKSHVNAVIGITPMNPYAGRSNLPLNKKLIFPSVAMSK
 PKDEIAEYMLYSQFNOAKQLSKHTVPFDQSEKLSQAQHYDFGLRAKLSVLSVG
 GLKRLALLETGDAESLGPDDVEPELIVQSIRETAPLLIKSDVEIMMEIESVCPGV
 MNTYPSOIVLVATVQVTDALADGCKDLQLLDFREVOVRLMADVLGDLEVL
 RKCKEOLITECVHORDVIEKLVKANSNTHYMWILLQRYVYNPGDFLQRLIKMAN
 AKNLGFEVLGVPDRVLRTPDRCLFTLITQALCQRLGSGPYGPAKGTISXKALGL
 QLGRITFCCDDTDNAGRIFIG

[illegible]

KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;
AUTHORS Eutheria: Primates: Catarrhini: Homidae: Homo.
1 (bases 1 to 139166)
Muzny,D., Arenson,A.D., Ding,Y., Dugan,S., Durbin,J., Forcum,J.,
Chacko,J., Chen,J., Garcia,C., Goodman,M., Gorrell,J.H., Haywood,M.,
Ganesh,R., Jackson,L., Jin,S., Kampal,R., Karpathy,S., Kovar,C.,
Hernandez,J., Lee,E., Li,Y., Lichtarge,O., Liu,W., Logan,O.,
Lau,S., Leat,B., Lee,E., Li,Y., Lichtarge,O., Liu,W., Logan,O.,
Lu,J., Ly,T., Marondel,I., Martinez,C., Verscher,S., Montgomery,K.,
Oswal,G., Perez,L., Rashid,N.D., Renault,B., Rowland,K., Savage,L.,
Scherer,S.E., Shen,H., Simon,M., Stovall,K., Timms,K.M., Todd,J.,
Vo,Q., Williamson,A., Worley,K.C., Yu,W., Kucherlapati,R.,
Nelson,D. and Gibbs,R.A.

• Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 139166)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (27-OCT-1997) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 139166)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (31-JUL-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 4 (bases 1 to 139166)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (16-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
REFERENCE 5 (bases 1 to 139166)
AUTHORS Worley,K.C.
TITLE Direct Submission
JOURNAL Submitted (17-SEP-1998) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT On Jul 31, 1998 this sequence version replaced gi:2909690.
Sequencing is completed to a minimum standard of double strand
coverage with a minimum of 2 clones and 2 reads with no ambiguities
or 2 chemistries with a minimum of 2 clones and 3 reads with no
ambiguities. If the sequence quality does not meet this standard,
it will be indicated in the annotation.

The repeat regions shown were identified using RepeatMasker by
Adrian Smit.

Sequence similarities were identified using Powerblast by Jinghui
Zhang.

Exon/Intron boundaries of identified genes were chosen if there
were canonical splice junctions that maintained sequence continuity
across the splice junctions.

FEATURES
source 1. 139166
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RPC13-462E2"
/chromosome="12q24"
1. 226
/note="Overlaps with bases 1-226 in AC002996."
/note="Region: Overlap with AC002996"
gene join(88..201,843..945,1380..1564,6860..7024,8688..8845,
12641..12755,18325..18759)
/gene="Human nucleus-encoded mitochondrial aldehyde
dehydrogenase(ALDH)gene"
/note="Unigene cluster Hs.74630, K03001, X05409"
530..696
repeat_region /rpt_family="AluJb"

repeat_region 1648..1732
/rpt_family="MIR"
complement(2001..2165)
/rpt_family="FRAM"
complement(2260..2370)
/rpt_family="LIMC3"
complement(2374..2612)
/rpt_family="AluSx"
complement(2631..2799)
/rpt_family="LIMC2"
complement(2800..3099)
/rpt_family="AluSc"
complement(3100..3220)
/rpt_family="LIMC2"
complement(3291..3588)
/rpt_family="AluJo"
complement(3608..4064)
/rpt_family="LIMB7"
complement(4069..4364)
/rpt_family="AluSx"
complement(4374..4672)
/rpt_family="AluSg"
4761..4893
/rpt_family="FLAM_C"
4904..5116
/rpt_family="AluSg/x"
5403..5684
/rpt_family="AluJb"
complement(5744..6014)
/rpt_family="AluSg"
6185..6402
/rpt_family="AluSg/x"
complement(7325..7487)
/rpt_family="FRAM"
7507..7790
/rpt_family="AluSx"
complement(7800..7943)
/rpt_family="AluSg"
complement(7947..8248)
/rpt_family="AluSx"
complement(8255..8389)
/rpt_family="FLAM_C"
complement(9202..9433)
/rpt_family="AluJo"
9452..9785
/rpt_family="LIMB6"
complement(9788..10134)
/rpt_family="THE1B"
10141..10197
/rpt_family="LIME1"
10201..10493
/rpt_family="AluSx"
10495..10754
/rpt_family="LIMB6"
10755..10916
/rpt_family="AluSc"
10918..11213
/rpt_family="AluSg"
11214..11379
/rpt_family="AluSc"
11381..11681
/rpt_family="AluY"
11693..11988
/rpt_family="AluSg"
11996..12094
/rpt_family="(TAAAA)n"
12100..12229
/rpt_family="AluJo"
12241..12529
/rpt_family="AluSx"
complement(12530..12557)
/rpt_family="AT-rich"
12631..12765
STS

```
/standard_name="DWC10/DWC11"
/note="GDB:179565, Chr. 12q24.2-1-2q24.2, Homo Sapiens"
13248. .13548
/rpt_family="AluX"
repeat_region complement(13595. .13622)
/rpt_family="AT-rich"
repeat_region 13999. .14103
/rpt_family="AluJo"
repeat_region complement(14148. .14404)
/rpt_family="AluB"
repeat_region 14525. .14811
/rpt_family="AluSc"
repeat_region 14814. .15118
/rpt_family="AluY"
repeat_region 15472. .15768
/rpt_family="AluSx"
repeat_region 15767. .16064
/rpt_family="AluY"
repeat_region complement(16333. .16541)
/rpt_family="L2"
repeat_region complement(16558. .16956)
/rpt_family="AluSx"
repeat_region 16992. .17186
/rpt_family="L1MB8"
repeat_region 17205. .17506
/rpt_family="AluSx"
repeat_region complement(17550. .17928)
/rpt_family="L2"
repeat_region 17954. .18213
/rpt_family="AluJo"
STS 18511. .18642
/gene="Human nucleus-encoded mitochondrial aldehyde
dehydrogenase(ALDH)gene"
/standard_name="SHGC-11012"
/note="D12S1884, RH637, STS17970, GDB:674606,
GDB:675353, Chr. 12pter-12qter between D12S84 and D12S369,
Homo Sapiens"
/db_xref="dbSTS:17970"
18807. .19105
/rpt_family="AluSx"
repeat_region complement(19388. .19650)
/rpt_family="AluJo"
repeat_region 19837. .20151
/rpt_family="AluSp"
repeat_region complement(20254. .20555)
/rpt_family="AluSx"
repeat_region complement(20559. .21050)
/rpt_family="LTR13"

Query Match 66.7%; Score 18; DB 11; Length 139166;
Best Local Similarity 80.8%; Pred. NO. 1.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAGCTGTTCTACCACTGAACCTAGG 26
||||| ||||| ||| ||||| |||
Db 100730 ATGCATGTTCTACTACATGATCTAGG 100705

RESULT 13
LOCUS HS856G1/c 151795 bp DNA HTG 09-MAY-1999
DEFINITION Homo sapiens chromosome 6 clone 856G1, WORKING DRAFT SEQUENCE, in
unordered pieces.
ACCESSION AL033381
NID 94468331
VERSION AL033381.2 GI:4468331
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 151795)
AUTHORS Phillips,S.
```

```
TITLE Direct Submission
JOURNAL Submitted (09-MAY-1999) Wellcome Trust Genome Campus, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humqueresanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Mar 22, 1999 this sequence version replaced gi:4376009.
IMPORTANT: This sequence is unfinished and does not necessarily
represent the correct sequence. Work on the sequence is in progress
and the release of this data is based on the understanding that the
sequence may change as work continues. The sequence may be
contaminated with foreign sequence from E.coli, yeast, vector,
phage etc. Order of segments is not known; 800 n's separate
segments. Unfinished: dj856G1 Contig_ID: 01260 acc=AL033381
Length: 151795 bp.
* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Location/Qualifiers
1..151795
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/clone="856G1"

BASE COUNT 41447 a 33248 c 34218 g 42882 t
ORIGIN

Query Match 66.7%; Score 18; DB 34; Length 151795;
Best Local Similarity 80.8%; Pred. NO. 1.5e+02;
Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 1 AAGCTGTTCTACCACTGAACCTAGG 26
||||| ||||| ||| ||||| |||
Db 17688 AAGCTGTTCTCCCAAGTGAACCTGG 17663

RESULT 14
AC004340
LOCUS AC004340 51050 bp DNA INV 07-JUL-1998
DEFINITION Drosophila melanogaster DNA sequence (Pl DS07020 (D238)), complete
sequence.
AC004340
NID g3293206
VERSION AC004340.1 GI:3293206
KEYWORDS HTG.
SOURCE Drosophila melanogaster (Subclones in tet from Pl clone DS07020
(D238)) DNA.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 51050)
AUTHORS Celniker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Agbayani,A., Arcaina,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Hummasti,S.R., Karra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S.,
Pfeiffer,B., Punch.E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
Sequencing of Drosophila chromosome, region 22E1-22E1
Unpublished (1997)
REFERENCE 2 (bases 1 to 51050)
AUTHORS Celniker,S.E., George,R.A., Galle,R., Svirskas,R.R., Hoskins,R.A.,
Agbayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G., Chavez,C.,
Chew,M., Doyle,C.M., Farfan,D.E., Flanagan,J., Houston,K.A.,
Hummasti,S.R., Karra,K., Kearney,L., Kim,S.H., Lee,B.,
Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S., Moshrefi,A.R.,
Moshrefi,M., Nixon,K., Pacleb,J.M., Park,S., Pfeiffer,B., Punch.E.,
Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R., Yee,A., Zhang,R.,
Zieran,L.L. and Kimmel,B.E.
Direct Submission
Submitted (07-MAR-1998) Drosophila Genome Center, Lawrence Berkeley
```

Laboratory, MS 64-121, Berkeley, CA 94720, USA
 On Jul 7, 1998 this sequence version replaced gi:3228456.
 Sequence submitted by:
 Berkeley Drosophila Genome Project
 Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our location
 archive web site (<http://fruitfly.berkeley.edu/sequence/>) or send
 email to drosophila@hgsc.lbl.gov.
 Library location: 12-74
 Location/Qualifiers
 source
 1. 51050
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2L"
 /map="22E1-22E1"
 /clone="P1 DS07020 (D238)"
 BASE COUNT 14368 a 11189 c 11359 g 14134 t
 ORIGIN
 Query Match 66.7%; Score 18; DB 36; Length 51050;
 Best Local Similarity 80.8%; Pred. No. 1.5e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 AGCTTGTTCTACCACTGAAGTGGC 27
 ||||| || ||||| ||||| |||||
 Db 9746 AACTGGTCCACCACGACGACGAGC 9771
 RESULT 15
 AC001657/c
 LOCUS
 DEFINITION Drosophila melanogaster DNA sequence (P1 DS06874 (D19)), complete
 sequence.
 ACCESSION AC001657 L36286 L36305 L36304 L36289 L36300 L39666 L36301 L36298
 L81427 L36303 L81426 L39670 L39668 L36294 L36285 L39667
 L36296 L36287 L36299 L36291 L36297 L36284 L39671 L39665 L39669
 L36292 L36293 L36302 L36290
 Q2337899
 NID AC001657.1 GI:2337899
 VERSION
 KEYWORDS
 SOURCE HTG.
 ORGANISM Drosophila melanogaster (Subclones in sac from P1 clone DS06874
 (D19)) DNA.
 Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 85139)
 Celniker, S.E., George, R.A., Galle, R.F., Hoskins, R.A.,
 Svirskaas, R.R., Harris, N.L., Agbayani, A., Arcaina, T.T., Baxter, E.,
 Blazej, R.G., Chavez, C., Chew, M., Doyle, C.M., Farfan, D.E.,
 Flanagan, J., Houston, K.A., Hummasti, S.R., Karra, K., Kearney, L.,
 Kim, S.H., Lee, B., Lomotan, M.A., Mak, J., Mazda, P., Mok, M.S.,
 Moshrefi, A.R., Moshrefi, M., Nixon, K., Pacleb, J.M., Park, S.,
 Pfeiffer, B., Punch, E., Snir, E., Twomey, B., Wan, K.H., Whitelaw, K.R.,
 Yee, A., Zhang, R., Zierar, L.L. and Kimmel, B.
 Sequencing of Alcohol dehydrogenase region
 * Unpublished (1997)
 TITLE
 JOURNAL
 REFERENCE 2 (bases 1 to 85139)
 AUTHORS Martin, C.H., Arcaina, T.T., Bondoc, M.M., Chiang, A., Critz, P.A.,
 Davis, C.A., Doyle, C.M., Ericsson, C.L., Farfan, D.E., Gunning, K.M.,
 Houston, K.A., Jaklevic, M.A., Radner, K.E., Kim, K., Kim, S.F.,
 Ko, C.L., Lewis, K.D., Li, M., Lindquist, K.J., Lomotan, M.A.,
 Lustre, V.M., Nachrus, M.U., Mayeda, C.A., Miguel, T.M., Miller, C.A.,
 Mok, M.S., Pacleb, J.M., Patel, S.G., Santos, R.F., Subramanian, S.,
 Wan, K.H., Whitelaw, K.R., Yee, A., Yeh, R.T., Yu, C. and Palazzolo, M.J.
 Direct Submission
 Submitted (22-APR-1997)
 TITLE
 JOURNAL
 COMMENT On Aug 20, 1997 this sequence version replaced gi:1945585.
 Sequence submitted by:
 Berkeley Drosophila Genome Project

Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our location
 archive web site (<http://fruitfly.berkeley.edu/sequence/>) or send
 email to drosophila@hgsc.lbl.gov.
 Library location: 58-72.
 Location/Qualifiers
 source
 1. 85139
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /chromosome="2L"
 /map="35B7-35C1"
 /clone="P1 DS06874 (D19)"
 /note="This sequence has not changed since its original
 submission on 04/22/1997. It was resubmitted in order to
 include all secondary accession numbers for the subclones
 belonging to this clone."
 BASE COUNT 25995 a 16491 c 16856 g 25797 t
 ORIGIN
 Query Match 66.7%; Score 18; DB 37; Length 85139;
 Best Local Similarity 80.8%; Pred. No. 1.5e+02;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 1 AAGCTTGTTCTACCACTGAAGTGG 26
 ||||| || ||||| ||||| |||||
 Db 41560 AAGCTGTTTATACCATCTGAACGAG 41535
 Search completed: September 18, 1999, 15:48:25
 Job time: 1385 sec

GenCore version 4.5

Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 18, 1999, 04:59:15 ; Search time 213.04 Seconds
(without alignments)
31.709 Million cell updates/sec

Title: US-09-037-472-1

Perfect score: 27

Sequence: 1 AAGCTGTGTTCTACCACTGAACTAGGC 27

Scoring table: IDENTITY_NUC

Searched: 311585 seqs, 125096042 residues

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	27	100.0	27	V32389	Interleukin-1-alpha
2	22	81.5	22	T70316	Primer for detecti
3	22	81.5	5701	X02988	Human IL-1ra BAC c
4	21	77.8	21	X16611	Interleukin 1 (441
5	18	66.7	327	V78683	Staphylococcus aur
6	17.4	64.4	633	V75647	Staphylococcus aur
7	17.2	63.7	687	T92382	Bacillus smithii n
8	17.2	63.7	648	V77989	Staphylococcus aur
9	17	63.0	1253	N40312	Sequence of human
10	17	63.0	6422	X20576	Polynucleotide seq
11	16.8	62.2	8160	X13096	Enterococcus faeca
12	16.6	61.5	7753	Q28727	Encodes exons XV t
13	16.6	61.5	59	V79270	Staphylococcus aur
14	16.6	61.5	400	V78290	Staphylococcus aur
15	16.6	61.5	361	V78258	Staphylococcus aur
16	16.6	61.5	400	V78053	Staphylococcus aur
17	16.6	61.5	400	V78055	Staphylococcus aur
18	16.6	61.5	401	V78042	Staphylococcus aur
19	16.6	61.5	553	V78001	Staphylococcus aur
20	16.6	61.5	400	V77971	Staphylococcus aur
21	16.6	61.5	400	V77942	Staphylococcus aur
22	16.6	61.5	475	V77932	Staphylococcus aur
23	16.6	61.5	1171	V77850	Staphylococcus aur
24	16.6	61.5	400	V77853	Staphylococcus aur
25	16.6	61.5	6591	V77425	Staphylococcus aur
26	16.6	61.5	237	V78877	Staphylococcus aur
27	16.6	61.5	239	V78855	Staphylococcus aur
28	16.6	61.5	239	V78854	Staphylococcus aur
29	16.6	61.5	411	V78664	Staphylococcus aur
30	16.6	61.5	339	V78574	Staphylococcus aur
31	16.6	61.5	400	V78545	Staphylococcus aur
32	16.6	61.5	381	V78334	Staphylococcus aur
33	16.6	61.5	400	V78340	Staphylococcus aur
34	16.6	61.5	1421	X13384	Enterococcus faeca
35	16.4	60.7	1393	Q36747	Sequence of TB-1 w
36	16.4	60.7	2259	Q46782	40 kDa and 34 kDa
37	16.2	60.0	384	T21919	Human gene signatu
38	16.2	60.0	1557	T70441	H. pylori AlpB gen
39	16.2	60.0	1557	V07248	DNA encoding Helic
40	16.2	60.0	4190	V34854	Human retinal dege
41	16.2	60.0	1922	V84426	Human secreted pro
42	16.2	60.0	1838	V80605	Kidney injury asso
43	16.2	60.0	110000	X20248_03	Continuation (4 of

c 44 16.2 60.0 5504 1 X20559 Polynucleotide seq
c 45 16.2 60.0 73 1 X30293 Borrelia burgdorfe

ALIGNMENTS

RESULT 1

V32389

ID V32389 standard; DNA; 27 BP.

AC V32389;

DT 11-SEP-1998 (first entry)

DE Interleukin-1-alpha primer 1 (-967/-945).

KW IL-1-alpha; genetic polymorphism; PCR; primer; amplification;

KW sight threatening diabetic retinopathy; interleukin-1-alpha;

KW interleukin-1-beta; interleukin-1RN; ss.

OS Synthetic.

OS Homo sapiens.

PN WO9815653-A1.

PD 16-APR-1998.

PF 09-OCT-1997; G02790.

PR 10-OCT-1996; GB-021129.

PA (DUFF/) DUFF G.

PA (RENN/) RENNIE I.

PA (RICH/) RICHARDSON R.

PI Duff G, Rennie I, Richardson R;

DR WPI: 98-240835/21.

PT Predicting increased risk of sight-threatening diabetic retinopathy
PT - comprises identifying genetic polymorphism pattern for genes
IL-1A, IL-1B and IL-1RN, useful to allow treatment before clinical

PT symptoms occur

PS Claim 2; Page 33; 41pp; English.

CC Interleukin-1-alpha (IL-1-alpha) primers 1 and 2 (V32390) were used to

CC amplify the IL-1-alpha gene region to identify single base variation

CC polymorphism of C/T at base 889. The invention claims to provide a

CC method for predicting the risk of sight threatening diabetic retinopathy.

CC The method involves isolating DNA from a patient and determining the DNA

CC polymorphism pattern of the genes that code for interleukin-1-alpha,

CC interleukin-1-beta and interleukin-1RN. The polymorphic pattern

CC identified is then compared with controls of known DNA polymorphism

CC patterns thereby identifying patients carrying a genetic polymorphism

CC associated with increased risk of sight threatening diabetic retinopathy.

CC The method may be able to identify diabetic patients at risk before the

CC determination of IL genes involved PCR reactions using primers V32389-

CC V32398. The method is also claimed to be useful in conjunction with

CC identification of other genes associated with sight threatening diabetic

CC retinopathy in genomic DNA and therefore, in identifying diabetic

CC patients expressing multiple risk patterns. 7 A; 8 C; 5 G; 7 T;

SQ Sequence 27 BP; 7 A; 8 C; 5 G; 7 T;

Query Match

Best Local Similarity 100.0%; Score 27; DB 1; Length 27;

Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AAGCTGTGTTCTACCACTGAACTAGGC 27

DB 1 AAGCTGTGTTCTACCACTGAACTAGGC 27

RESULT 2

T70316

ID T70316 standard; DNA; 22 BP.

AC T70316;

DT 19-OCT-1997 (first entry)

DE Primer for detecting genetic predisposition to periodontal disease.

KW Periodontal disease; gingivitis; periodontitis; polymorphism;

KW interleukin-1 alpha; IL-1A; interleukin-1 beta; IL-1B; primer; PCR;

KW polymerase chain reaction; ss.

OS Synthetic.

PN WO9706180-A1.

PD 20-FEB-1997.

PF 02-AUG-1996; U12455.
PR 03-AUG-1995; US-510696.
PA (KORN/-) KORNMAN K S.
PI (MEDI-) MEDICAL SCI SYSTEMS INC.
PI Duff GW, Kornman KS;
DR WPI: 97-154207/14.
PT Identification of patient's genetic polymorphism pattern - allows
PT prediction of increased periodontal disease severity
PS Claim 3; Page 25; 33pp; English.
CC PCR primers (T70316 and T13884-88) can be used in a new method
CC for identifying a genetic predisposition to periodontal disease by
CC detecting the presence of DNA polymorphisms in the gene sequences
CC for interleukin-1 alpha (IL-1A) and interleukin-1 beta (IL-1B).
CC Alleles associated with severe disease were identified as IL-1A
CC allele 2 together with IL-1B (Tag1) allele 2. The primers are used
CC to amplify DNA from a blood or tissue sample and products are
CC subjected to restriction digestion to determine the polymorphism
CC pattern of the patient. A single base variation (C/T) polymorphism
CC at IL-1A base -889 can be identified using primers (T70316 and
CC T13884) corresponding to bases -967 to -945 and -888 to -869,
CC respectively. An NcoI site is created if C is available at -889,
CC but not if T is present.
SQ Sequence 22 BP; 5 A; 7 C; 4 G; 6 T;

Query Match 81.5%; Score 22; DB 1; Length 22;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 TGTCTTACCACCTGAACCTAGGC 27
DB 1 TGTCTTACCACCTGAACCTAGGC 22

RESULT 3
X02988
ID X02988 standard; DNA; 5701 BP.
AC X02988;
DT 22-JUN-1999 (first entry)
DE Human IL-1ra BAC contiguous DNA sequence 33.
KW Tango-77; human; IL-1ra; cytokine superfamily; inflammation; inhibition;
KW interleukin-1 receptor; IL-1R; regulation; asthma; rheumatoid arthritis;
KW chronic myelogenous leukaemia; psoriasis; inflammatory bowel disease;
KW growth factors; treatment; IL-1 receptor complex; BAC; ss.
OS Homo sapiens.
PN WO9906426-A1.
PD 11-FEB-1999.
PF 03-AUG-1998; U16102.
PR 02-JUL-1998; US-091650.
PA 04-AUG-1997; US-054646.
PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PI Pan Y;
DR WPI: 99-153692/13.
PT New isolated nucleic acid encoding the new human cytokine Tango-77 -
PT used to inhibit inflammation and to screen for specific modulators
PS Example 5; Figure 3; 226pp; English.
CC X02956-X03048 and X22301-X22304 are overlapping BAC genomic sequences
CC containing alternatively spliced forms of human IL-1ra. Such fragments
CC are used in the method of the invention which describes the isolation of
CC a novel human TANGO-77 encoding nucleic acid and protein. Tango-77 is a
CC member of the cytokine superfamily that is expected to inhibit
CC inflammation by binding to the interleukin-1 receptor (IL-1R). It may
CC also bind to a new receptor so could regulate other cellular processes
CC associated with acute or chronic inflammation, e.g. asthma, chronic
CC myelogenous leukaemia, rheumatoid arthritis, psoriasis and inflammatory
CC bowel disease. It may also induce or suppress interleukins, cytokines and
CC growth factors. Modulators of this protein are used to treat or prevent
CC conditions associated with abnormal levels of inflammation, or activity
CC of IL-1 or its receptor complex.
SQ Sequence 5701 BP; 1729 A; 1238 C; 1079 G; 1654 T;

Query Match 81.5%; Score 22; DB 1; Length 5701;
Best Local Similarity 100.0%; Pred. No. 0.12;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Query Match

81.5%; Score 22; DB 1; Length 5701;

Best Local Similarity

Matches

22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 5

V78683

ID V78683 standard; DNA; 327 BP.

Query Match

77.8%; Score 21; DB 1; Length 21;

Best Local Similarity

Matches 21; Conservative 0; Mismatches 0; Indels 0;

V78683;
 16-MAR-1999 (first entry)
 Staphylococcus aureus contig SEQ ID #4372.
 Computer readable medium; vaccine; S.aureus infection; immunodetection;
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 skin infection; surgical wound infection; scalded skin syndrome;
 toxic shock syndrome; ds.
 Staphylococcus aureus.
 EP-786519-A2.
 30-JUL-1997.
 07-JAN-1997; 100117.
 05-JAN-1996; US-009861.
 (HUMA-) HUMAN GENOME SCI INC.
 Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI; 97-374922/35.
 Polynucleotide(s) and proteins derived from Staphylococcus aureus
 stored on computer readable medium and used in the production of
 anti-S.aureus vaccines
 Claim 1: Page 2938; 3271pp; English.
 This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 of the invention. The DNA sequences are recorded on a computer readable
 medium, preferably selected from a floppy or hard disk, random access
 memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 the S.aureus DNA sequences allows putative functions to be assigned so
 that protein-encoding or regulatory regions of commercial, therapeutic or
 industrial importance can be obtained. Specifically, sequences which are
 likely to encode antigens have been identified and these polypeptides can
 be used in a vaccine composition against S.aureus infection. The
 S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 skin and surgical wound infections, scalded skin syndrome, toxic shock
 syndrome, etc. Organisms transformed with the DNA sequences can be used
 for recombinant production of the polypeptides. The new DNA sequences
 (and their fragments) are useful as primers or probes for isolating
 homologues of any of the S.aureus DNA sequences contained on the
 computer readable medium.
 Sequence 327 BP; 87 A; 85 C; 75 G; 77 T;
 Query Match 66.7%; Score 18; DB 1; Length 327;
 Best Local Similarity 80.8%; Pred. No. 12;
 Matches 21; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
 QY 2 AGCTTGTCTTACCACCTGACTAGGC 27
 || ||| ||||| ||| |||||
 Db 170 AGTTGCTCTACCAATTGACGTAGGC 195
 RESULT 6
 V75647/c
 ID V75647 standard; DNA; 633 BP.
 AC V75647;
 DT 16-MAR-1999 (first entry)
 DE Staphylococcus aureus contig SEQ ID #1336.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 skin infection; surgical wound infection; scalded skin syndrome;
 toxic shock syndrome; ds.
 Staphylococcus aureus.
 EP-786519-A2.
 30-JUL-1997.
 07-JAN-1997; 100117.
 05-JAN-1996; US-009861.
 (HUMA-) HUMAN GENOME SCI INC.
 Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 DR WPI; 97-374922/35.
 Polynucleotide(s) and proteins derived from Staphylococcus aureus
 stored on computer readable medium and used in the production of
 anti-S.aureus vaccines
 Claim 1: Page 1924; 3271pp; English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative functions to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 633 BP; 252 A; 105 C; 90 G; 184 T;

Query Match 64.4%; Score 17.4; DB 1; Length 633;
 Best Local Similarity 94.7%; Pred. No. 25;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 5 TTGTTCTACCACCTGAACCT 23
 |||| ||||| ||||| |||||
 Db 86 TTGCTGCTACCACCTGAACCT 68

RESULT 7
 T92382/c
 ID T92382 standard; DNA; 687 BP.
 AC T92382;
 DT 48-JAN-1998 (first entry)
 DE Bacillus smithii nitrile hydratase subunit beta encoding DNA.
 KW acrylonitrile; acrylamide; nitrile hydratase subunit beta;
 KW thermally stable protein; ds.
 OS Bacillus smithii.
 FH Key Location/Qualifiers
 FT CDS 1..687
 FT /*tag= a
 FT /product= Nitrile_hydratase_subunit_beta
 FT /note= "No stop codon given"
 FT
 PN J09248188-A.
 PD 22-SEP-1997.
 PF 18-MAR-1996; 060732.
 PR 18-MAR-1996; JP-060732.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI; 97-520742/48.
 DR P-PSDB; W32621.
 PT Gene encoding nitrile hydratase - for producing amide from nitrile
 PT by biological catalysis
 PS Claim 7; Page 11-12; 10pp; Japanese.
 CC The present sequence encodes nitrile hydratase subunit beta, a novel
 CC protein isolated from Bacillus smithii. The protein has hydration
 CC activity for converting acrylonitrile into acrylamide. It is useful
 CC for producing amide from nitrile by biological catalysis.
 SQ Sequence 687 BP; 212 A; 112 C; 173 G; 190 T;

Query Match 63.7%; Score 17.2; DB 1; Length 687;
 Best Local Similarity 86.4%; Pred. No. 32;
 Matches 19; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 GCTTTGTTCTACCACCTGAACCTA 24
 |||| ||||| ||||| |||||
 Db 374 GCTTTTCTACAACCTTAACCTA 353

RESULT 8
 V77989

```

ID AC V77989 standard; DNA: 648 BP.
DE 16-MAR-1999 (first entry)
KW Staphylococcus aureus contig SEQ ID #3678.
KW Computer readable medium; vaccine; S aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 121..180
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification for this DNA sequence."
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997. 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI: 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 2639-2640; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 648 BP; 161 A; 150 C; 144 G; 131 T;

Query Match 63.7%; Score 17.2; DB 1; Length 648;
Best Local Similarity 82.6%; Pred. No. 31;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TTGTTTACCACCTCACTAGGC 27
||| ||||| ||| |||||
DB 9 TTGNTTACCAATTGAGTAGGC 31

RESULT 9
N40312
ID N40312 standard; DNA: 1253 BP.
AC N40312, 1992 (first entry)
DE Sequence of human leukocyte interferon (leIFN) gene.
KW Interferon; yeast expression vector; promoter; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 52..621
FT /*tag= a
FT W08401153-A.
PN 29-MAR-1984.
PD 09-SEP-1983; U01370.
PF 15-SEP-1982; US-418521.

```

```

PA (COLB ) COLLABORATIVE RES INC.
PI Botstein D, Bowden DW, Davis RW, Fink GR, Knowlton RG,
PI Mao J, Taunton-Rigby A, Vovis GF;
DR WPI: 84-088509/14.
DR P-PSDB; P40741.
PT Prodn. of interferon in yeast - by use of recombinant DNA segment
PT contg. suc 2 promoter linked to interferon gene
PS Example; Page 48; 53pp; English.
CC The inventors claim a method for the prodn. of interferon in yeast
CC by use of recombinant DNA segment contg. suc 2 promoter linked to
CC interferon gene. The interferon synthesis can be regulated over a
CC broad range by changes in the culture medium, e.g. synthesis is
CC repressed in glucose medium and de-repressed when glucose is
CC depleted. Also claimed is yeast strain ATCC 20644-CGY144.
SQ Sequence 1253 BP; 373 A; 268 C; 244 G; 368 T;

Query Match 63.0%; Score 17; DB 1; Length 1253;
Best Local Similarity 80.0%; Pred. No. 42;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTTACCACCTGAACCTAG 25
||| ||| ||| ||||| ||
DB 2 AAGCTTGTCTCATCATCTGTGAACCA 26

RESULT 10
X20576
ID X20576 standard; DNA: 6422 BP.
AC X20576;
DT 05-MAY-1999 (first entry)
DE Polynucleotide sequence from the genome of Treponema pallidum.
KW Treponema pallidum infection; syphilis; Borrelia infection; animal;
KW enzyme production; ds.
OS Treponema pallidum.
PN W09859034-A2.
PD 30-DEC-1998.
PF 23-JUN-1998; U13041.
PR 24-JUN-1997; US-050667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Fraser CM;
DR WPI: 99-081273/07.
PT New isolated Treponema pallidum nucleic acids - used to develop
PT products for the detection, diagnosis, characterisation, prevention
PT and therapy of T. pallidum infections, particularly syphilis
PS Claim 1: Page 593-597; 1150pp; English.
CC X20500-21243 represent polynucleotide sequences from the genome of
CC Treponema pallidum. The sequences can be used for detection,
CC diagnosis, characterisation, prevention and therapy for T. pallidum
CC infections, particularly syphilis. They can also be used for detecting
CC diseases related to Borrelia infections in animals, and for the
CC production of biosynthetic products such as enzymes.
SQ Sequence 6422 BP; 1614 A; 1888 C; 1547 G; 1361 T;

Query Match 63.0%; Score 17; DB 1; Length 6422;
Best Local Similarity 80.0%; Pred. No. 53;
Matches 20; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 3 GCTTGTCTTACCACCTGAACCTAGGC 27
||| ||||| ||| |||||
DB 1198 GCGTGTCTTACCAACTGAGCTATGC 1222

RESULT 11
X13096
ID X13096 standard; DNA: 8160 BP.
AC X13096;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:159.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KW vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.

```

```
PN WO9850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; US-08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1; Page 898-902; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 8160 BP; 2731 A; 1463 C; 1654 G; 2292 T;

Query Match 62.2%; Score 16.8; DB 1; Length 8160;
Best Local Similarity 90.0%; Pred. No. 68;
Matches 18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 6 TGTCTACCACTGACTAG 25
DB 3366 TGTCTAGCTCTGACTAG 3385

RESULT 12
Q26727/c
ID Q26727 standard; DNA; 7753 BP.
AC Q26727;
DE Encodes exons XV to XVIII of human hepatocyte growth factor.
KW HGF; enhance growth; preparing transgenic animals; hepatic disease;
KW clinical diagnostic reagent; drug.
OS Homo sapiens.
FH Key Location/Qualifiers
FT cds 101..241
FT /tag= a
FT /note= "exon XV"
FT cds 773..879
FT /tag= b
FT /note= "exon XVI"
FT cds 991..1136
FT /tag= c
FT /note= "exon XVII"
FT cds 3764..3937
FT /tag= d
FT /note= "exon XVIII"
FT poly_a_signal 7492..7497
FT /tag= e

J04183394-A.
PD 30-JUN-1992.
PF 19-NOV-1990; 314548.
PR 19-NOV-1990; JP-314548.
PA (NAKA/) NAKAMURA T.
PA (TOYM ) TOYOBO KK.
DR WPI; 92-265591/32.
DR P-PSDB; R25692.
PT Recombinant human hepatocyte growth factor and DNA encoding it -
```

```
PT useful for diagnosis and treatment of hepatic disease and
PT transgenic animal prepn.
PS Disclosure; Page 22; 28pp; Japanese.
CC This sequence contains exons XV to XVIII of human hepatocyte growth
CC factor. See also R25676-92, Q26713-27.
SQ Sequence 7753 BP; 2567 A; 1177 C; 1243 G; 2766 T;

Query Match 61.5%; Score 16.6; DB 1; Length 7753;
Best Local Similarity 82.6%; Pred. No. 84;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AGCTGTCTTACCCTGAACTA 24
DB 6273 ACCCTGTCTTACCCTAAATA 6251

RESULT 13
V79270
ID V79270 standard; DNA; 59 BP.
AC V79270;
DE 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #4959.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus -
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1; Page 3132; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 59 BP; 16 A; 18 C; 11 G; 14 T;

Query Match 61.5%; Score 16.6; DB 1; Length 59;
Best Local Similarity 82.6%; Pred. No. 43;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TTGTTCTACCACTGACTAGGC 27
DB 35 TTGCTCTACCAATTGAGCTAGGC 57

RESULT 14
V78290
ID V78290 standard; DNA; 400 BP.
```


AC V78290;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #3979.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
FH Key Location/Qualifiers
FT misc_feature 121..180
FT /*tag= a
FT /note= "these bases represent a line of missing text in
FT the sequence listing in the specification. They
FT are included to maintain the nucleotide numbering
FT given in the specification. for this DNA sequence".
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 2772-2773; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 400 BP; 87 A; 90 C; 84 G; 75 T;

Query Match 61.5%; Score 16.6; DB 1; Length 400;
Best Local Similarity 82.6%; Pred. No. 56;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TTGTTCTACCACTGAGCTAGGC 27
DB 294 TTGCTCTACCAATTGAGCTAGGC 316

RESULT 15
V78258
ID V78258 standard; DNA; 361 BP.
AC V78258;
DT 16-MAR-1999 (first entry)
DE Staphylococcus aureus contig SEQ ID #3947.
KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
OS Staphylococcus aureus.
PN EP-786519-A2.
PD 30-JUL-1997.
PF 07-JAN-1997; 100117.
PR 05-JAN-1996; US-009861.
PA (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
PI Rosen CA;
DR WPI; 97-374922/35.
PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
PT stored on computer readable medium and used in the production of
PT anti-S.aureus vaccines
PS Claim 1: Page 2759; 3271pp; English.
CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences
CC of the invention. The DNA sequences are recorded on a computer readable
CC medium, preferably selected from a floppy or hard disk, random access
CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
CC the S.aureus DNA sequences allows putative functions to be assigned so
CC that protein-encoding or regulatory regions of commercial, therapeutic or
CC industrial importance can be obtained. Specifically, sequences which are
CC likely to encode antigens have been identified and these polypeptides can
CC be used in a vaccine composition against S.aureus infection. The
CC polypeptides can also be used in a kit for the immunodetection of
CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
CC skin and surgical wound infections, scalded skin syndrome, toxic shock
CC syndrome, etc. Organisms transformed with the DNA sequences can be used
CC for recombinant production of the polypeptides. The new DNA sequences
CC (and their fragments) are useful as primers or probes for isolating
CC homologues of any of the S.aureus DNA sequences contained on the
CC computer readable medium.
SQ Sequence 361 BP; 104 A; 83 C; 78 G; 93 T;

Query Match 61.5%; Score 16.6; DB 1; Length 361;
Best Local Similarity 82.6%; Pred. No. 55;
Matches 19; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 5 TTGTTCTACCACTGAGCTAGGC 27
DB 47 TTGCTCTACCAATTGAGCTAGGC 69

Search completed: September 18, 1999, 05:27:13
Job time: 1678 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic acid nucleic search, using sw model

Run on: September 18, 1999, 05:46:43 ; Search time 1405 Seconds
(without alignments)
37.906 Million cell updates/sec

Title: US-09-037-472-1

Perfect score: 27
Sequence: 1 AAGCTTGTTCTACCCACCTGAAGTAGGC 27

Scoring table: IDENTITY_NUC

Searched: 2546578 seqs, 986266752 residues

Database :

EST:

1: em_est1:*
2: em_est2:*
3: em_est3:*
4: em_est4:*
5: em_est5:*
6: em_est6:*
7: em_est7:*
8: em_est8:*
9: em_est9:*
10: em_est10:*
11: em_est11:*
12: em_est12:*
13: em_est13:*
14: em_est14:*
15: em_est15:*
16: em_est16:*
17: em_est17:*
18: em_est18:*
19: em_est19:*
20: gb_est1:*
21: gb_est2:*
22: gb_est3:*
23: gb_est4:*
24: gb_est5:*
25: gb_est6:*
26: gb_est7:*
27: gb_est8:*
28: gb_est9:*
29: gb_est10:*
30: gb_est11:*
31: gb_est12:*
32: gb_est13:*
33: gb_est14:*
34: gb_est15:*
35: gb_est16:*
36: gb_est17:*
37: gb_est18:*
38: gb_est19:*
39: gb_est20:*
40: gb_est21:*
41: gb_est22:*
42: gb_est23:*
43: gb_est24:*
44: gb_est25:*
45: gb_est26:*
46: gb_est27:*
47: gb_est28:*
48: gb_est29:*
49: gb_est30:*
50: gb_est31:*
51: gb_est32:*
52: em_est20:*
53: em_est21:*

54: em_est22:*
55: em_est23:*
56: em_est24:*
57: em_est25:*
58: em_est26:*

Pred. No. is the number of results predicted by chance +0 have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	18.6	68.9	405	25	N88530	N88530 K3598F Huma
C 2	18.6	68.9	525	36	AA607268	AA607268 vm92h05.f
C 3	18.6	68.9	459	37	AA681861	AA681861 vt44a05.s
C 4	18	66.7	224	32	AA331881	AA331881 EST35751
C 5	18	66.7	564	33	AA422885	AA422885 vda47c06.s
C 6	18	66.7	422	39	AA824084	AA824084 vt66c03.s
C 7	17.8	65.9	465	43	AA116163	AA116163 A004P750
C 8	17.6	65.2	337	20	T05736	T05736 EST03625 Fe
C 9	17.6	65.2	308	21	F13213	F13213 HSC3KC111 n
C 10	17.6	65.2	526	21	T60600	T60600 yb91g05.r1
C 11	17.6	65.2	425	27	AA004244	AA004244 zh88d06.r
C 12	17.6	65.2	480	27	AA004527	AA004527 zh91a09.r
C 13	17.6	65.2	593	48	AA156606	AA156606 tn52h04.x
C 14	17.6	65.2	226	50	AV046966	AV046966 AV046966
C 15	17.4	64.4	399	20	T10245	T10245 seq1305 b4H
C 16	17.4	64.4	308	20	T32622	T32622 EST1965 Hu
C 17	17.4	64.4	294	20	T32523	T32523 EST81965 Hu
C 18	17.4	64.4	267	20	T35508	T35508 EST86615 Hu
C 19	17.4	64.4	323	20	T35551	T35551 EST87186 Hu
C 20	17.4	64.4	281	20	T35556	T35556 EST87349 Hu
C 21	17.4	64.4	425	21	R17871	R17871 Y910d02.r1
C 22	17.4	64.4	383	21	R18749	R18749 Y917c08.r1
C 23	17.4	64.4	484	22	R54291	R54291 Y97f06.r1
C 24	17.4	64.4	442	24	H94074	H94074 Yw58a10.r1
C 25	17.4	64.4	273	31	AA300763	AA300763 EST13619
C 26	17.4	64.4	556	31	AA306372	AA306372 EST177333
C 27	17.4	64.4	549	31	AA306551	AA306551 EST177501
C 28	17.4	64.4	395	31	AA307033	AA307033 EST177948
C 29	17.4	64.4	503	31	AA307495	AA307495 EST178379
C 30	17.4	64.4	594	31	AA308049	AA308049 EST178890
C 31	17.4	64.4	571	31	AA310973	AA310973 EST181747
C 32	17.4	64.4	528	31	AA311686	AA311686 EST182410
C 33	17.4	64.4	401	31	AA314612	AA314612 EST186444
C 34	17.4	64.4	354	31	AA316439	AA316439 EST188137
C 35	17.4	64.4	407	47	AI508192	AI508192 mh31e12.y
C 36	17.4	64.4	815	47	AI525561	AI525561 PT1.3.03
C 37	17.4	64.4	670	49	C96306	C96306 C96306 Marc
C 38	17.4	64.4	459	49	C96449	C96449 C96449 Marc
C 39	17.4	64.4	708	53	HSM001010	HSM001010 Homo sapi
C 40	17.2	63.7	280	20	D39825	D39825 RICS1442A R
C 41	17.2	63.7	452	38	AA808387	AA808387 ca90d09.s
C 42	17.2	63.7	443	39	AA860156	AA860156 ak47b09.s
C 43	17.2	63.7	238	48	AI558929	AI558929 fb78b02.y
C 44	17	63.0	334	23	H51722	H51722 yp85e07.r1
C 45	17	63.0	502	54	HSM010701	HSM010701 Homo sapi

ALIGNMENTS

RESULT 1
N88530/c
LOCUS N88530 405 bp mRNA
DEFINITION K3598F Human fetal heart, Lambda ZAP Express Homo sapiens CDNA
clone K3598 5', mRNA sequence.
ACCESSION N88530
NID G1441732
VERSION N88530.1 GI:1441732

EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 REFERENCE 1 (bases 1 to 405)
 AUTHORS Liew,C.C.
 TITLE cDNAs from fetal heart (1996)
 JOURNAL Unpublished (1996)
 COMMENT On Sep 1, 1995 this sequence version replaced.

Contact: Liew CC
 Department of Laboratory Medicine and Path. biology
 University of Toronto
 Banting Institute, 100 College St., Toronto, Ontario, M5G1L5
 Tel: 4169788758
 Fax: 4169785650
 Email: liewcc@utcc.utoronto.ca
 Seq primer: GAAATTAACCTCACTAAGGG.

FEATURES
 source
 1..405
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="K3598"
 /clone.lib="Human fetal heart, Lambda ZAP Express"
 /lab.host="E. coli XL1-Blue"
 /note="Vector: Lambda ZAP Express; Site_1: EcoRI; Site_2:
 XhoI; mRNA was purified from human fetal hearts (8-10
 weeks). cDNA was synthesized using a XhoI-Oligo dT
 adaptor-primer. EcoRI adaptors were ligated, followed by
 digestion with XhoI, for directional cloning into
 predigested lambda ZAP Express."

BASE COUNT 100 a 112 c 113 g 80 t
 ORIGIN

Query Match 68.9%; Score 18.6; DB 25; Length 405;
 Best Local Similarity 84.0%; Pred. No. 52;
 Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0

QY 3 GCTTGTCTTACCACCTGAACCTAGGC 27
 ||||| ||| |||| |||| |||| ||||
 Db 59 GCTTGTCTTGCACCTGAACCAAGGC 35

RESULT 2
 AA607268/c
 LOCUS
 DEFINITION
 ACCESSION
 NID
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 525)
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1288177.

Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800

```

Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:569969
High quality sequence stop: 472.
Location/Qualifiers
  1..525
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:1005753"
/clone_lib="Knowles Solter mouse blastocyst Bl"
/tissue_type="blastocyst"
/dev_stage="embryo (pre-implantation)"
/lab_host="DH10B"
/note="Organ: embryo; Vector: pSPORT; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally from mRNA prepared
from 800 blastocysts. Primer: SalI{dl};
5'-CGCTCGACCGTCACCGTCTTTTCTTTT-3'. cDNAs were
cloned into the NotI/SalI sites of a pSPORT vector (Life
Technologies). Two different size selections: Bl (larger
inserts) and B3."
BASE COUNT      132 a      129 c      111 g      153 t
ORIGIN

Query Match      58.9%; Score 18.6; DB 36; Length 525;
Best Local Similarity 84.0%; Pred. No. 55;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTACCACTGAACCTAG 25
|||||
Db 288 AAGCTTGTCTCCACCTCACCTAG 264
|||||

RESULT 3
AA681861/c
LOCUS
DEFINITION
ACCESSION
NID
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kubaca, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1392950.

Contact:: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicinep
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:612808
High quality sequence stop: 432.
Location/Qualifiers
  1..459
/organism="Mus musculus"

```

```

/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/clone="IMAGE:1123472"
/clone_lib="Knowles Solter mouse 2 cell"
/tissue_type="embryo"
/dev_stage="2-cell"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pBluescribe (modified);
Site_1: MluI; Site_2: SalI; Cloned unidirectionally from
mRNA prepared from 13,500 2-cell stage embryos. Primer:
Sali(dT): 5'-CGGTCGACCGTCGACGTTTGTTTT-3'. CDNAS
were cloned into the MluI/Sali sites of a modified
pBluescribe vector using commercial linkers (NEB).
Average insert size: 1.2 Kb."
BASE COUNT 105 a 112 c 101 g 141 t
ORIGIN

Query Match 68.9%; Score 18.6; DB 37; Length 459;
Best Local Similarity 84.0%; Pred. No. 54;
Matches 21; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Oy 1 AAGCTTGTTCTACCACTGAAGTAG 25
||||||| ||||| |||||
Db 167 AAGCTTGTTCTCCCACTCACTAG 143

RESULT 4
AA331881/c
LOCUS AA331881 224 bp mRNA EST 21-APR-1997
DEFINITION EST35751 Embryo, 8 week I Homo sapiens cDNA 5' end similar to
similar to MER5 protein, mRNA sequence.
ACCESSION AA331881
NID 91984123
VERSION AA331881.1 GI:1984123
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 224)
Adams,M.D., Kerlavage,A.R., Fleischmann,R.D., Fuldner,R.A.,
Bult,C.J., Lee,N.H., Kirkness,E.F., Weinstock,K.G., Gocayne,J.D.,
White,O., Sutton,G., Blake,J.A., Brandon,R.C., Man-Wai,C.,
Clayton,R.A., Cline,T.R., Cotton,M.D., Earle-Hughes,J., Fine,L.D.,
Fitzgerald,L.M., Fitzhugh,W.M., Fritchman,J.L., Geoghagen,N.S.,
Glodek,A., Gnehm,C.L., Hanna,M.C., Hedblom,E., Hinkle,P.S.Jr.,
Kelley,J.M., Kelley,J.C., Liu,L.-I., Marmaros,S.M., Merrick,J.M.,
Moreno-Palauques,R.F., McDonald,L.A., Nguyen,D.T., Pelligrino,S.M.,
Phillips,C.A., Ryder,S.E., Scott,J.L., Saudek,D.M., Shirley,R.,
Small,K.V., Spriggs,T.A., Utterback,T.R., Weidman,J.F., Li,Y.,
Badnarik,D.P., Cao,L., Cepeda,M.A., Coleman,T.A., Collins,E.J.,
Dimke,D., Feng,D.-F., Ferrie,A., Fischer,C., Hastings,G.A.,
He,W.W., Hu,J.S., Greene,J.M., Gruber,J., Hudson,P., Kim,A.K.,
Kozak,D.L., Kunsch,C., Hungjun,J., Li,H., Weissner,P.S., Olsen,H.,
Raymond,L., Wei,Y.F., Wing,J., Xu,C., Yu,G.L., Ruben,S.M.,
Dillon,P.J., Fannon,M.R., Rosen,C.A., Haseitine,W.A., Fields,C.,
Fraser,C.M. and Venter,J.C.
Initial assessment of human gene diversity and expression patterns
based upon 83 million nucleotides of cDNA sequence
JOURNAL Nature 377 (6547 Suppl), 3-174 (1995)
MEDLINE 96026280
COMMENT On Sep 12, 1996 this sequence version replaced gi:1405042.
Other_ESTs: THCL70844
Contact: Kerlavage, AR
Bioinformatics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850 USA
Tel: 3018699056
Fax: 3018699423
Email: arkerlavet@tigr.org
For clone availability, additional sequence and expression
information related to this EST, please check the TIGR Human Gene

```

```

Index (http://www.tigr.org/tdb/hgi/hgi.html)
Seq primer: M13 Reverse.
Location/Qualifiers
1..224
/organism="Homo sapiens"
/db_xref="ATCC (inhost):133506"
/clone_lib="Embryo, 8 week I"
/dev_stage="embryo, 8 wks"
/notes="Organ: Embryo, 8 weeks; Vector: pBluescript SK-;
Site_1: EcoRI; Site_2: XhoI"

```

```

BASE COUNT 42 a 60 c 60 g 56 t
ORIGIN

```

```

Query Match 66.7%; Score 18; DB 32; Length 224;
Best Local Similarity 77.8%; Pred. No. 88;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Oy 1 AAGCTTGTTCTACCACTGAAGTAGGC 27
||||||| ||||| ||||| |||||
Db 149 AAGCTGTTCTCCACATGCAGNAGGC 123

```

```

RESULT 5
AA422885
LOCUS AA422885 564 bp mRNA EST 16-OCT-1997
DEFINITION vd47c06.sl Knowles Solter mouse 2 cell Mus musculus cDNA clone
IMAGE:803722 5' similar to gb:X64550 M.musculus mRNA for
hyaluronan-mediated motility (MOUSE);, mRNA sequence.
ACCESSION AA422885
NID 92101727
VERSION AA422885.1 GI:2101727
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 564)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
COMMENT On May 8, 1995 this sequence version replaced gi:800408.
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
MG1:48066
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 441.
Location/Qualifiers
1..564
/organism="Mus musculus"
/strain="B6D2 F1/J"
/db_xref="taxon:10090"
/map="15"
/clone="IMAGE:803722"
/clone_lib="Knowles Solter mouse 2 cell"
/dev_stage="embryo"
/tissue_type="embryo"
/lab_host="DH10B"
/notes="Organ: embryo; Vector: pBluescribe (modified);
Site_1: MluI; Site_2: Sali; Cloned unidirectionally from

```

```

FEATURES
source

```


4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

Insert Size: 1027
 High quality sequence stops: 283 Source: IMAGE Consortium, LLNL This
 clone is available royalty-free through LLNL ; contact the IMAGE
 Consortium (info@image.llnl.gov) for further information.
 Insert Length: 1027 Std Error: 0.00
 Seq primer: M13RP1
 High quality sequence stop: 283.

FEATURES

Location/Qualifiers
 1..526
 /organism="Homo sapiens"
 /db_xref="GDB:498329"
 /db_xref="taxon:9606"
 /clone="IMAGE:78584"
 /clone_lib="Stratagene liver (#937224)"
 /sex="male"
 /dev_stage="49 years old"
 /lab_host="SOLR cells (kanamycin resistant)"
 /note="Organ: liver; Vector: pBluescript SK; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dT. Hepatectomy from normal male caucasian. Average
 insert size: 1.1 kb; Uni-ZAP XR Vector; ~5' adaptor
 sequence: 5' GAATCGGCACGAG 3' ~3' adaptor sequence: 5'
 CTCGAGTTTTTTTTTTTTTTT 3' "

BASE COUNT 142 a 117 c 126 g 136 t 5 others
 ORIGIN

Query Match 65.2%; Score 17.6; DB 21; Length 526;
 Best Local Similarity 83.3%; Pred. No. 1.6e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTACACCTGAACCTA 24
 |||||
 Db 279 AAGCTTGTCTACACCTGAACCTA 302

RESULT 11

AA004244/c
 LOCUS
 DEFINITION
 zh88d06.r1 Soares fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
 clone IMAGE:428363 5' similar to contains element MER22 repetitive
 element ;, mRNA sequence.

ACCESSION
 NID
 VERSION
 KEYWORDS
 SOURCE

AA004244.1 GI:11447919
 EST.
 human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Wards, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
 and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

97044478

COMMENT

On May 18, 1995 this sequence version replaced gi:811621.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 Insert Length: 673 Std Error: 0.00
 Seq primer: mob.REGA-E1
 High quality sequence stop: 395.

FEATURES

Location/Qualifiers
 1..425
 /organism="Homo sapiens"
 /db_xref="GDB:1328132"
 /db_xref="taxon:9606"
 /clone="IMAGE:428363"
 /clone_lib="Soares_fetal_liver_spleen_INFLS_S1"
 /sex="male"
 /dev_stage="20 week-post conception fetus"
 /lab_host="DH10B (ampicillin resistant)"
 /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
 with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
 This is a subtracted version of the original Soares fetal
 liver spleen INFLS library. 1st strand cDNA was primed
 with a Pac I - oligo(dT) primer [5',
 AACTGGAACAATTAATTAAAGATCTTTTTTTTTTTTTTTT 3'],
 double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Pac I and cloned into the Pac I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization. Library
 constructed by Bento Soares and M. Patricia Bonaldo."

BASE COUNT 100 a 105 c 97 g 120 t 3 others
 ORIGIN

Query Match 65.2%; Score 17.6; DB 27; Length 425;
 Best Local Similarity 83.3%; Pred. No. 1.5e+02;
 Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTACACCTGAACCTA 24
 |||||
 Db 83 AAGCTTGTCTACACCTGAACCTA 60

RESULT 12

AA004527/c
 LOCUS
 DEFINITION
 zh91a09.r1 Soares fetal_liver_spleen_INFLS_S1 Homo sapiens cDNA
 clone IMAGE:428632 5' similar to contains Alu repetitive
 element; contains element MER22 repetitive element ;, mRNA sequence.

ACCESSION
 NID
 VERSION
 KEYWORDS
 SOURCE

AA004527.1 GI:1448104
 EST.
 human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

Hillier, L., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,
 Chissoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,
 Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,
 Wards, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,
 Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,
 Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,
 and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags
 Genome Res. 6 (9), 807-828 (1996)

97044478

COMMENT

On May 9, 1995 this sequence version replaced gi:802992.

Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the

source

further details.

FEATURES
source

Location/Qualifiers
1. .226
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/map="14q24.3"
/clone="1700065B04"
/clone_lib="Mus musculus adult C57BL/6J testis"
/sex="male"
/tissue_type="adult"
/dev_stage="testis"

BASE COUNT 55 a 58 c 38 g 75 t
ORIGIN

Query Match 65.2%; Score 17.6; DB 50; Length 226;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 20; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 2 AGCTTGTTCTACCACTGAAGTAG 25
|||||
Db 154 AGTTTTTCTACCTCTGAAGTAG 177

RESULT 15

T10245/c

LOCUS T10245 399 bp mRNA EST 07-JUN-1994
DEFINITION seq1305 b4HB3MA Cot8-HAP-Ft Homo sapiens cDNA clone
b4HB3MA-COT8-HAP-Ft213 5' similar to similar to Mouse MER5 protein,
mRNA sequence.

ACCESSION T10245

NID 9471594

VERSION T10245.1 GI:471594

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 399)
Soares,M.B., Bonaldo,M.F., Jelenc,P., Su,L., Lawton,L and
Estratiadis,A.
Construction and characterization of a normalized cDNA library
Proc. Natl. Acad. Sci. U.S.A. 91, 9228-9232 (1994)
95023884

COMMENT

Contact: Bento Soares
Columbia University
Department of Psychiatry, 722 W 168th Street, Unit #41, New York, NY
10032
Tel: 2129602313
Fax: 2127813577
Email: cuc@cucfa.ccc.columbia.edu
sense, Ampicillin
Seq primer: M13 Reverse.
Location/Qualifiers

FEATURES
source

1. .399
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="b4HB3MA-COT8-HAP-Ft213"
/clone_lib="b4HB3MA Cot8-HAP-Ft"
/lab_host="E. coli"
/note="Vector: Lfamid BA; Site.1: Not I; Site.2: Hind III;
Size-selected cDNA from polyA+ RNA from human brain.
3-month old neonate. This library is the result of an
attempted normalization of library b4HB3MA. "
90 a 93 c 97 g 119 t

BASE COUNT
ORIGIN

Query Match 64.4%; Score 17.4; DB 20; Length 399;
Best Local Similarity 77.8%; Pred. No. 1.9e+02;
Matches 21; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 AAGCTTGTCTTACCACCTGAAGTAGGC 27
|||||
Db 116 AAGCTCGTTCTTCCACATGCAGCAGGC 90

Search completed: September 18, 1999, 06:47:28
Job time: 3645 sec